DIGITALIA

RIVISTA DEL DIGITALE NEI BENI CULTURALI

Anno XX, Numero 2 - **2025**

ICCU

MODUS

ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane
www.iccu.sbn.it

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista
è ammessa con obbligo di citazione

«DigItalia», rivista del digitale nei beni culturali, è una rivista *peer reviewed*
e segue il codice etico delle pubblicazioni

I contributi possono essere proposti alla rivista dagli autori o su iniziativa del Comitato Scientifico.
Gli articoli vengono sottoposti al Comitato di Redazione per un primo accertamento sulla corrispondenza con i campi di ricerca della rivista. I contributi delle sezioni Saggi e Progetti vengono indirizzati in forma anonima ad almeno uno studioso di comprovata competenza sui temi affrontati. I revisori fanno pervenire i loro giudizi alla redazione (favorevole alla pubblicazione, favorevole con modifiche/miglioramenti, non favorevole). Se il giudizio finale è positivo, viene comunicata agli autori l'accettazione del contributo, insieme ad eventuali indicazioni suggerite dai valutatori, di cui si garantisce comunque l'anonimato.

DIGITALIA

Rivista del digitale nei beni culturali
ISSN 1972-621X
Anno XX, Numero 2 - Dicembre 2025

In copertina:
L'immagine è una libera elaborazione grafica della testa della statua di Apollo del I sec. d.c. (Civitavecchia, Museo Nazionale), copia da un originale greco avvicinabile all'Apollo di Leochares (IV sec. a.c.)

Direttore Fondatore
Marco Paoli

Direttore Responsabile
Giuliano Genetasio

Comitato di Redazione
Caporedattrice
Valentina Menesatti

Amalia Maria Amendola
Valentina Atturo
Laura Borsi
Elisabetta Castro
Silvana de Capua
Maria Cristina Mataloni
Marianna Morreale
Lucia Negrini
Federica Olivotto
Claudia Pecoraro
Angelo Restaino
Elisa Sciotti
Alice Semboloni
Maria Lucia Violo

Grafica & Impaginazione
MLA&Partner - Roberta Micchi

Produzione e Stampa
Istituto Poligrafico e Zecca dello Stato S.p.A.
Roma

Editore
ICCU
Istituto centrale per il catalogo unico
delle biblioteche italiane
Viale Castro Pretorio, 105
00185 Roma
T +39 06 49.210.425
email: ic-cu.digitalia@cultura.gov.it
https://digitalia.cultura.gov.it

ICCU

Comitato Scientifico

Stefano Allegrezza
Enrico Pio Ardolino
Emanuele Bellini
Giovanni Bergamin
Julian Bogdani
Dimitri Brunetti
Flavia Bruni
Simonetta Buttò
Rossella Caffo
Elisabetta Caldelli
Gianfranco Crupi
Andrea De Pasquale
Cristina Dondi
Pierluigi Feliciati
Giuliano Genetasio
Marina Giannetto
Klaus Kempf
Nicolette Mandarano
Maurizio Messina
Laura Moro
Maria Teresa Natale
Marco Paoli
Don Valerio Pennasso
Marco Pizzo
Roberto Raieli
Gino Roncaglia
Maria Letizia Sebastiani
Giovanni Solimine
Laura Tallandini
Anna Maria Tammaro
Antonella Trombone
Valentina Vassallo
Chiara Veninata

SOMMARIO

dicembre 2025

DIGITALIA 2005-2025
Testimonianze, riflessioni e prospettive

Un passato che è già futuro: vent'anni di DigItalia 9
di Giuliano Genetasio

Nascita di una rivista 11
di Marco Paoli

Il digitale: il contesto europeo, Europeana e le iniziative italiane 15
di Rossella Caffo

Una rivista per il dialogo interdisciplinare 22
di Simonetta Buttò

SAGGI

L'impatto dell'European Accessibility Act sul settore culturale: obblighi e opportunità per archivi, musei e biblioteche 29
di Stefano Allegrezza

Il museo nell'era digitale: tecnologie sensibili per una umanità aumentata 45
di Claudia Pecoraro

Vantaggi e limiti della modellazione solida applicata all'archeologia 61
di Silvana Rapuano

PROGETTI

Fondi, Possessori ed Esemplari nell'Indice SBN 81
di Maria Cristina Mataloni, Alice Semboloni

Early Modern Digital Itineraries: The Italian Core Set 97
di Rachel Midura

Le Digital Humanities alla Biblioteca nazionale universitaria di Torino: sul progetto BINARI e dintorni 111
di Alessia Grillone

The watermark portal Bernstein-Memory of Paper and its Italian contributions 127
di Emanuel Wenger

Transkribus e la trascrizione automatica di manoscritti. Un caso studio sulle carte di Vittorio Alfieri 139
di Sara Gallegati

Visualising Interpretive Layers in Cultural Heritage: Data narratives with Linked Open Data in mythLOD collection 151
di Valentina Pasqual, Francesca Tomasi

Cultural Heritage Accessibility through Multisensory Interactions: Updates from the MuseIT project 161
di Nasrine Olson, Flavia Massara, Vasilis Papadopoulos, Maria Kyrou, Albert Meroño-Peñuela, Moa Johansson, Vyacheslav Tykhonov, James Hanlon

DOCUMENTI E DISCUSSIONI

Alfabeto arabo e OCR open source: un'analisi grafica e linguistica dei risultati di elaborazione per il recupero di dati bibliografici 179
di Riccardo Amerigo Vigliermo

Invented archives. Un problema teoretico per la ricerca storica o per l'archivistica? 203
di Salvatore Spina

EVENTI E SEGNALAZIONI

Biblioteche e trasformazione digitale: un percorso formativo e-learning a cura della Scuola nazionale del patrimonio e delle attività culturali e dell'Associazione Italiana Biblioteche 229
di Giovanni Bergamin, Gaia Caltavituro, Marzia Piccininno, Andrea Razza

Europeana Climate Action Community and the path to regenerative digital transformation 236
di Evangelia Paschalidou

DIGITALIA 2005-2025
Testimonianze, riflessioni e prospettive

MODUS

Un passato che è già futuro: vent'anni di DigItalia

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00136

Giuliano Genetasio
Direttore dell'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU)

La fine del 2025 segna il ventennale di «DigItalia. Rivista del digitale nei beni culturali», una ricorrenza che rappresenta una tappa significativa nella storia di una pubblicazione ormai consolidata nel panorama scientifico nazionale e internazionale.
Sin dagli esordi DigItalia si è distinta per una spiccata proiezione verso il futuro, facendo del precoce interesse per le tecnologie digitali e della trasversalità culturale secondo la prospettiva MAB (biblioteche, archivi, musei) – con grande anticipo rispetto alla nascita di un coordinamento ufficiale in Italia – cifre essenziali del proprio discorso, che rendono la rivista, ancora oggi, così attuale.

Confrontandomi con la redazione ho deciso di celebrare questa ricorrenza con alcuni interventi sul passato e sul presente della rivista. L'intento è stato quello di lasciare il giusto spazio alla componente scientifica – che è e resta il cuore pulsante della rivista – senza tuttavia rinunciare a sottolineare un anniversario così emblematico.

Ho voluto che a scrivere questi interventi fosse direttamente chi ha diretto la rivista negli anni passati dalla sua fondazione agli anni più recenti, dando così voce alle visioni di chi ha plasmato con le proprie mani l'evoluzione della rivista.

Marco Paoli, fondatore della rivista nel 2005 e direttore fino al 2008, ricostruirà il contesto che ha portato alla nascita di DigItalia. Il suo contributo evidenzierà le motivazioni, gli obiettivi iniziali e il legame con i numerosi progetti di digitalizzazione che, in quegli anni, prendevano forma sia a livello nazionale sia europeo.

Rosa Caffo, direttrice della rivista dal 2009 al 2014, offrirà una riflessione sul progetto di una Biblioteca Digitale Europea – divenuto poi Europeana – e sui principali programmi di digitalizzazione avviati in quegli anni, parlandoci poi della piattaforma MOVIO, sviluppata dall'ICCU per la realizzazione di mostre virtuali e ancora oggi largamente utilizzata.

Simonetta Buttò, direttrice dal 2015 a metà 2024, ripercorrerà gli sviluppi più recenti: dall'adozione della piattaforma OJS a partire dal 2012, all'accreditamento di DigItalia presso l'ANVUR per diversi settori scientifico-disciplinari. Il suo intervento toccherà anche il percorso che ha condotto, a partire dal 2019, alla realizzazione del Sistema integrato di ricerca, poi confluito in Alphabetica, uno degli sviluppi più rilevanti degli ultimi anni.

MODUS

E arriviamo così al presente. Ho assunto la direzione di DigItalia a partire da luglio 2024. In questi primi mesi ho avviato un profondo rinnovamento del Comitato scientifico, con un duplice obiettivo: rafforzarne da un lato la dimensione internazionale – un percorso che è tuttora in corso – e, dall'altro, ampliare il ventaglio degli ambiti disciplinari rappresentati, riequilibrando la storica prevalenza delle aree bibliotecaria e archivistica, per valorizzare maggiormente le altre discipline riconosciute oggi dall'ANVUR nell'ambito della rivista.

Questa evoluzione della composizione del Comitato scientifico si è a sua volta riflessa nei contributi scientifici della rivista, sempre più provenienti dai vari settori disciplinari della cultura purché con un focus sul digitale.

Da questo numero è prevista inoltre una revisione del sito della rivista, che recepirà le più recenti direttive in materia di accessibilità e introdurrà il formato HTML, accanto a diversi aggiornamenti grafici, tra cui una nuova copertina.

Buona lettura, e buon compleanno DigItalia.

Nascita di una rivista

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00137

Marco Paoli
Già direttore dell'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU)

Risalire all'origine di un'idea non è sempre agevole, in specie quando il contesto storico e scientifico è stato determinante assai più delle motivazioni soggettive e personali. Queste ultime erano all'epoca piuttosto evidenti. Giungevo alla direzione dell'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU) nel gennaio del 2005, con alle spalle esperienze condotte in biblioteche di tradizione e di natura prevalentemente conservativa, dove l'approccio alla digitalizzazione era stato praticamente nullo. Avrei dovuto quindi calarmi nella nuova realtà, confrontandomi in primis con i progetti che in quei mesi si stavano attuando, e che risultavano innovativi per lo stesso Istituto. I termini della sfida erano proprio questi, e qui si annidava anche un potenziale rischio.

Le attività dell'ICCU erano numerose e rilevanti[1], ma ne spiccavano due, destinatarie di cospicui finanziamenti erogati nel 2003 dal Comitato dei Ministri per la Società dell'Informazione, e su cui l'Istituto aveva investito adeguate risorse umane: la costruzione del portale Internet Culturale e la partecipazione al progetto denominato Biblioteca Digitale Italiana (BDI), esteso a biblioteche statali e regionali allo scopo di coordinare le singole campagne di digitalizzazione nel settore dei fondi musicali, fotografici e cartografici. La prima release del network, su cui avrebbero dovuto confluire i contenuti digitali elaborati in seno alla BDI[2], veniva messa online nel marzo del 2005, dopo l'avvenuta presentazione ufficiale ai ministri Urbani e Stanca; e di lì a poco sarebbe partita la seconda release del portale con la realizzazione di un servizio di acquisizione a pagamento delle immagini ad alta risoluzione da parte degli utenti. Nel febbraio l'ICCU aveva inoltre ospitato un workshop sul tema della *digital preservation* con la partecipazione di un nutrito gruppo di esperti stranieri, e in quell'occasione era stato presentato il volume degli atti della conferenza internazionale di Firenze del 2003 sul Futuro delle memorie digitali e patrimonio culturale[3].

[1] Ne davo un sintetico bilancio in Marco Paoli, *Stato dell'arte e prospettive di lavoro dell'ICCU*, «Bollettino AIB», 45 (2005), n. 1, p. 85-92, <https://bollettino.aib.it/article/view/5392>.

[2] Se ne veda un primo resoconto in Marco Paoli, *I progetti di digitalizzazione della Biblioteca Digitale Italiana*, «DigItalia. Rivista del digitale nei beni culturali», (2005), n. 0, p. 85-91, <https://digitalia.cultura.gov.it/article/view/331>.

[3] *Futuro delle memorie digitali e patrimonio culturale: atti del Convegno internazionale, Firenze, 16-17 ottobre 2003*, a cura di V. Tola, C. Castellani, Roma: ICCU, 2004.

MODUS

In quel periodo, come è noto, il contesto europeo era altrettanto ricco di iniziative nell'ambito del digitale. Si era appena concluso il Progetto MINERVA (2002-2005), finanziato dalla Commissione europea, cui l'Italia aveva partecipato, finalizzato allo scambio di buone pratiche e alla definizione di linee guida, metadati e standard relativi alla digitalizzazione; e, nell'aprile dello stesso 2005, il Ministero per i beni e le attività culturali annunciava a Roma la partecipazione del nostro paese al progetto MICHAEL, anch'esso promosso dalla Commissione europea, con lo scopo di realizzare un portale multilingue per l'accesso alle risorse digitali culturali di Francia, Regno Unito e Italia. Oltreoceano, anche se il target non era quello culturale, iniziative di enorme rilievo stavano per decollare nel settore della digitalizzazione: nel gennaio 2005 Google lanciava la prima versione di Google Maps, e cinque mesi dopo era la volta del programma Google Earth.

Un momento, quindi, di indubbia aspettativa nei confronti delle moderne tecnologie e di fiducia nella possibilità delle amministrazioni pubbliche di dare risposte concrete ed efficaci alle esigenze che operatori culturali e utenti manifestavano da anni. L'ICCU, che procedeva a stretto contatto con la Direzione Generale di appartenenza, veniva ad essere al tempo stesso attivatore e destinatario del Nuovo, e la situazione fattuale sollecitava iniziative che apportassero sintesi e chiarezza, come, ad esempio, la creazione di una nuova rivista specializzata nel settore.

Si torna così alla questione iniziale, al recupero delle motivazioni e degli intenti che hanno determinato la nascita di «DigItalia. Rivista del digitale nei beni culturali». Vi era in proposito un altro elemento che non poteva essere trascurato. Ormai da alcuni anni, i due periodici dell'ICCU, «SBN Notizie», strumento di informazione sull'avanzamento del Servizio Bibliotecario Nazionale (SBN), e «Il Corsivo», notiziario del Censimento Nazionale delle Edizioni Italiane del XVI secolo (Edit16), erano cessati, il primo nel 2001, il secondo nel 1999. Era dunque avvertita l'esigenza di pubblicare nuovamente un organo dell'Istituto, e, in considerazione di quanto abbiamo fin qui esposto, ciò sarebbe dovuto avvenire, senza soverchie esitazioni, nell'ambito del digitale. Personalmente, si sarebbe trattato di ripetere - in un contesto completamente diverso, ma ugualmente stimolante - l'esperienza che mi aveva portato a fondare nel 1994, nel settore degli studi sul libro antico, la rivista «Rara Volumina»[4].

Ecco pronta per essere sottoposta alla valutazione dei collaboratori, l'idea di dar vita ad un periodico che ospitasse contributi italiani e stranieri sullo stato dell'arte della digitalizzazione nei beni culturali, e che promuovesse lo scambio di informazioni e proposte sui vari piani del dibattito, ora tecnico, ora gestionale, ora giuridico. L'accoglienza della proposta fu immediata (e lo fu anche il titolo della rivista che mi permisi di formulare, con la grafica particolare che metteva in risalto il nome del nostro paese), e si procedette in tempi brevi alla definizione della struttura di ogni numero, prevedendo, oltre ai saggi,

[4] https://it.wikipedia.org/wiki/Rara_volumina.

MODUS

sezioni per accogliere relazioni su progetti di digitalizzazione, resoconti su eventi significativi, documenti redatti da organismi nazionali e internazionali, ecc. Venne curata anche la veste editoriale, caratterizzata dalla scelta per la copertina di un bene culturale archeologico, la testa di una statua di Apollo conservata al Museo nazionale di Civitavecchia. Tengo a precisare che, in questa fase preliminare, il contributo della Vicedirettrice Anna Maria Mandillo fu determinante, e altrettanto significativo fu l'apporto dei responsabili delle varie aree, Gabriella Contardi, Claudia Leoncini, Cristina Magliano, Massimo Menna, e Giuliana Sgambati. Il Direttore Generale Luciano Scala e il Capo Dipartimento Salvatore Italia[5] giudicarono favorevolmente l'iniziativa e garantirono una propria presentazione del primo numero, come di fatto avvenne. Il Comitato scientifico fu accuratamente selezionato, fino a comprendere i responsabili della BDI, con Tullio Gregory e Amedeo Quondam, e i principali esperti di contenuti digitali italiani e stranieri, quali Anna Maria Tammaro, Costantino Thanos, Hans Hofman e Seamus Ross. La prima diffusione del periodico, sia cartacea che online, ebbe una risonanza nazionale[6] e internazionale[7], anche a seguito di una conferenza stampa organizzata presso l'ICCU.

Vennero da subito apprezzate le potenzialità del nuovo strumento informativo che contribuiva a diffondere e a far circolare esperienze teoriche, tecnologie, progetti, eventi, cornici istituzionali e normative, facendo emergere una pluralità di soggetti nazionali e internazionali, interessati a far conoscere risultati e prospettive, e destinatari a loro volta di preziose informazioni di ritorno. La strada era tracciata per i numeri successivi, e il panorama che si veniva delineando era dei più incoraggianti, con il risultato di accrescere le cognizioni specifiche di una comunità sempre più vasta e nel contempo di superare l'iperspecializzazione di pochi. Il comune denominatore della digitalizzazione era alla base dell'ampia adesione alla rivista da parte di contributori e lettori, ma costituiva anche la principale novità di essa. In un quadro di insieme accademico e scientifico in cui (ancora oggi e a maggior ragione venti anni fa) prevaleva la settorialità delle discipline, l'apertura di «DigItalia» ai beni culturali non solo librari e documentali ma anche archivistici, archeologici, storico-artistici, sonori e audiovisivi rappresentò certamente un passo in avanti verso l'interdisciplinarietà e l'integrazione delle conoscenze e delle metodologie. A livello ministeriale, un coordinamento tra i diversi programmi di digitalizzazione del patrimonio culturale è ora garantito, a far data dal 2020, dall'Istituto centrale per la digitalizzazione del patrimonio culturale (Digital Library), ed è quindi ormai acquisito ciò che sulle colonne della rivista, a partire dal 2005, è stato di fatto attuato, configurando un'assai più vasto dialogo, esteso a tematiche strategiche come la conservazione del

[5] Ricordo con simpatia che il prof. Italia, quando mi ricevette nel suo studio a Roma, non restò indifferente al fatto che il titolo della rivista, con la particolare grafica, riecheggiasse il suo cognome.

[6] *DigItalia: rivista del digitale nei beni culturali. N. 0 (dic. 2005),* [recensione di Alberto Petrucciani], «Bollettino AIB», 46 (2006), n. 1/2, p. 124-125 <https://bollettino.aib.it/article/view/5150>; Emilio Cabasino, recensione del numero zero in «Economia della Cultura», 16 (2006), n. 4, p. 589.

[7] Si veda il dettagliato resoconto del numero zero curato da Philippe Marcerou sul «Bulletin des bibliothèques de France» (2006), n. 6, p. 106-107, edito dall'Enssib, <https://bbf.enssib.fr/consulter/bbf-2006-06-0106-005>.

MODUS

web, le connessioni ipertestuali, la completezza dei metadati, gli standard di qualità nei progetti di digitalizzazione, e così via. Tale vocazione pluridisciplinare è ancora una delle caratteristiche di «DigItalia», ed è sufficiente in proposito menzionare, da ultimo, la relazione sul progetto Screnim, censimento dei graffiti carcerari d'età moderna[8], relativo ad una base dati di testimonianze incise sulle pareti di 150 ex prigioni italiane, graffiti ora figurativi, ora testuali: beni culturali di sicuro valore storico, quasi del tutto inaspettati.

Altra significativa caratteristica della rivista è l'attualità, intesa quale puntuale specchio di nuovi indirizzi operativi e di nuove tecnologie, ed è ciò che ne giustifica l'auspicabile prosecuzione negli anni a venire.

[8] Marco Albertoni, *Screnim: un progetto di censimento dei graffiti carcerari d'età moderna quali fonti storiche da conservare e valorizzare*, «DigItalia. Rivista del digitale nei beni culturali», 19 (2024), n. 22, p. 181-190, <https://digitalia.cultura.gov.it/article/view/3066>.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

Il digitale: il contesto europeo, Europeana e le iniziative italiane

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00138

Rossella Caffo
Già direttrice dell'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU)

DigItalia vede la luce nel 2005. Oggi si celebrano i vent'anni di questa rivista, edita dall'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU) e dedicata allo studio e al dibattito critico sull'applicazione delle tecnologie digitali ai vari settori del patrimonio culturale, allo scopo di offrire spazi agli approfondimenti e alle riflessioni sul tema. La storia di questi venti anni testimonia il processo per la creazione di una visione comune e condivisa tra gli Stati membri e tra questi e la Commissione europea, e le tante iniziative concrete per la realizzazione di biblioteche digitali, con l'obiettivo di favorire l'accesso in rete ai contenuti culturali. Per una singolare coincidenza, nello stesso anno in cui nasce DigItalia sei capi di Stato di paesi europei (Italia, Francia, Germania, Polonia, Spagna e Ungheria) indirizzano al presidente del Consiglio europeo e a quello della Commissione europea un messaggio che pone il tema della digitalizzazione del patrimonio culturale e della creazione di una Biblioteca digitale europea. La lettera dei capi di Stato dell'aprile 2005 suggeriva esplicitamente che la Commissione si impegnasse per creare in tempi brevi una Biblioteca digitale europea in grado di rendere accessibile a tutti i cittadini il patrimonio scientifico e culturale dell'Europa. Si trattava, come spiegava la lettera, di difendere un patrimonio «di una ricchezza e di una diversità che non ha eguali», che esprime «l'universalismo di un continente che, nel corso della sua lunga storia, ha dialogato con il resto del mondo». Il proposito era anche fermare l'avanzata di Google, che in sei anni si riproponeva di digitalizzare 15 milioni di libri (4,5 miliardi di pagine) provenienti da quattro prestigiose biblioteche americane e una inglese (Oxford), con l'intento di coinvolgere anche le biblioteche europee allo sviluppo del progetto Google books.

Ancor prima della lettera dei sei capi di Stato iniziative in questo campo erano state avviate con il supporto di alcuni governi e della Commissione europea.

Il 4 aprile 2001 rappresentanti ed esperti degli Stati membri si riunirono a Lund, in Svezia, per discutere della digitalizzazione del patrimonio culturale, con l'obiettivo di produrre raccomandazioni per il coordinamento dei programmi di digitalizzazione nazionali inquadrati nel contesto europeo.

La premessa di tale iniziativa risiedeva nel riconoscere e affermare che le risorse culturali e scientifiche europee costituiscono un patrimonio pubblico e unico su cui si basa la memoria collettiva delle nostre diverse società e lo sviluppo delle industrie di contenuti digitali per una società della conoscenza democratica e sostenibile.

I principi di Lund sono stati il primo importante documento ad affrontare un'analisi

MODUS

critica indicando alcune iniziative da assumere nel breve, medio e lungo periodo. Tra le criticità allora individuate c'erano la frammentazione dell'approccio alla digitalizzazione, la mancanza di standard condivisi, i problemi di copyright, la scarsa sinergia tra istituzioni culturali e i programmi di innovazione tecnologica, la conservazione del digitale, la poca attenzione al multilinguismo, gli scarsi investimenti e la mancanza di organici piani di sviluppo pluriennali sia ai vari livelli nazionali che a quello europeo.
Tra le proposte avanzate, emerse la creazione di un forum di cooperazione tra i vari Stati membri per elaborare una comune visione relativa a politiche e programmi, facilitare lo scambio di buone pratiche, definire indicatori per favorire processi di benchmarking, per creare punti di riferimento dedicati alle istituzioni culturali in campi quali i metadati, il multilinguismo, le tecnologie per la riproduzione delle immagini, la conservazione del digitale, gli standard tecnici per l'interoperabilità, con l'obiettivo di produrre raccomandazioni e linee guida utili all'intero processo di digitalizzazione e messa in linea dei contenuti culturali.
Su tali basi, pertanto, la Commissione, di concerto con gli Stati membri, istituiva il National Representatives Group (NRG), un gruppo di esperti nominati dagli Stati membri, che si riuniva ogni sei mesi sotto l'egida della Presidenza di turno e a cui la Commissione garantiva la segreteria. Il gruppo successivamente si rinnovò sulla base di decisioni della Commissione europea. Nel 2007 il nome con cui si identificava tale team era Member States Expert Group (MSEG), che dal 2017 cambia in Digital Cultural Heritage and Europeana (DCHE), concludendo i suoi lavori nel maggio 2021. La missione del NRG era monitorare lo stato dell'arte e i progressi relativi agli obiettivi contenuti nei principi di Lund, e produrre ogni sei mesi un report che registrasse la situazione, le iniziative e i progressi per ogni singolo Stato membro, che includesse anche raccomandazioni per azioni future condensate nel *Lund action plan*, aggiornato ogni anno. La Commissione provvedeva al sostegno operativo del lavoro del gruppo: nasceva così il progetto MINERVA, finanziato nell'ambito del Quinto Programma quadro di ricerca e sviluppo tecnologico e coordinato dal Ministero italiano[1].
Il progetto avviava e realizzava un'azione mirata al coordinamento delle politiche, dei programmi e dei progetti di digitalizzazione del patrimonio culturale tra i paesi membri dell'Unione europea. Gli obiettivi principali erano: rendere disponibile in rete l'immenso patrimonio europeo evitando frammentazioni, sprechi di risorse, duplicazioni degli sforzi, creare le condizioni favorevoli per lo scambio di esperienze e per definire una comune visione strategica per la creazione di uno spazio comune europeo dell'informazione culturale in rete.
MINERVA ha costituito il braccio operativo del NRG per l'implementazione dei principi di Lund e ha agito a supporto degli Stati membri nella preparazione del *National policy profile* e nella produzione del *Progress report* annuale che ha registrato e illustrato lo stato dell'arte, gli sviluppi e i progressi relativi alla digitalizzazione in tutti gli Stati membri.

[1] https://cordis.europa.eu/project/id/IST-2001-35461.

MODUS

Il progetto, avviato nel 2002, coinvolgeva i ministeri o gli enti governativi preposti alla cultura degli allora quindici paesi dell'Unione Europea. Nel 2004, con l'allargamento dell'Unione ad altri dieci Stati, la rete MINERVA, con un ulteriore finanziamento comunitario (progetto MINERVA Plus), si allargava a comprendere i nuovi paesi.
La rete MINERVA operava su due livelli, politico e tecnico. Il livello politico consisteva nel garantire una stretta collaborazione tra gli Stati membri attraverso i Ministeri titolari della competenza sul patrimonio culturale, e tra questi e la Commissione europea. In tale prospettiva MINERVA si proponeva inoltre di dare visibilità alle iniziative nazionali, di promuovere lo scambio di buone pratiche e di assicurare la diffusione e la conoscenza delle politiche e dei programmi comunitari a livello nazionale e locale.
Il livello tecnico riguardava invece la creazione di una comune piattaforma condivisa dagli Stati membri, costituita da raccomandazioni e linee guida per la digitalizzazione, che potessero favorire e garantire l'interoperabilità delle risorse digitalizzate, la qualità dei contenuti e dei servizi di accesso per la comunicazione e la fruizione del patrimonio culturale attraverso la rete, con un approccio integrato tra archivi, biblioteche e musei.
Contemporaneamente le Presidenze di turno dell'UE contribuivano a produrre una visione comune e condivisa relativa al patrimonio culturale che viene espressa e definita da documenti strategici quali ad esempio la *Carta di Parma* (2003), coordinata e pubblicata dalla Presidenza Italiana, mentre la Presidenza del Regno Unito ribadiva l'impegno e delineava un piano di azione europeo attraverso il *Dinamic Action Plan* (2004). I citati documenti vengono realizzati con il supporto del progetto MINERVA, ormai vera e propria parola chiave ben conosciuta in tutti gli stati membri e anche al di fuori dei paesi europei, come testimoniano altre iniziative, ad esempio il progetto MEDCULT[2].
In considerazione dei risultati raggiunti, nel 2006 viene approvato dalla Commissione europea il progetto MINERVA eC, terzo atto del percorso di MINERVA che consolidava e aggiornava quanto già realizzato e sviluppava azioni di supporto per la costruzione della Biblioteca digitale europea, in aderenza agli indirizzi comunitari[3].

[2] MEDCULT è un progetto finanziato dall'UNESCO che nel 2006, sotto il coordinamento del Ministero italiano, ha riguardato tre partner mediterranei, Egitto, Giordania e Marocco. Ha realizzato nei tre paesi altrettanti workshop rivolti a numerose istituzioni culturali, e basati sulla diffusione dei risultati di MINERVA, in particolare le linee guida per la realizzazione di siti web culturali accessibili e di qualità, tradotte in arabo nell'ambito del progetto.

[3] Sul sito <https://michael-culture.eu/library/> si possono trovare notizie sul progetto e le sue estensioni e i testi scaricabili dei documenti prodotti, oltre a quelli citati nel testo, tra i quali: *Technical guidelines for Digital Cultural Heritage creation programmes*, <https://michael-culture.eu/mca-library/minerva-technical-guidelines-for-digital-cultural-content-creation-programmes-version-2-0-2008/>; *Handbook on cultural web user interaction*, <https://michael-culture.eu/mca-library/handbook-on-cultural-web-user-interaction/>; *Guide for Intellectual property rights and other legal issues*, *Cost reduction in digitization,* <https://michael-culture.eu/mca-library/cost-reduction-in-digitisation/>; *Quality principles for cultural websites: a handbook*, <https://michael-culture.eu/mca-library/quality-principles-for-cultural-web-sites-a-handbook-2005/>; *Good practices handbook*, <https://michael-culture.eu/mca-library/minerva-good-practices-handbook/>. Inoltre all'indirizzo <https://web.archive.org/web/20240507091135/https://www.minervaeurope.org/>, si trova il sito di MINERVA in forma integrale dove è anche possibile scaricare tutti i documenti prodotti. Oscurato dal Ministero, il sito è stato salvato dall'iniziativa Internet Archive, che lo ha ritenuto di interesse per la conservazione.

Si inseriva pertanto in un processo già avviato, la citata lettera dei sei capi di Stato a cui la Commissione europea rispose con una svolta nella politica sulle biblioteche digitali. La strategia venne delineata nella Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, intitolata *i2010: le biblioteche digitali*, (inserita nel piano di Lisbona), pubblicata nel settembre 2005[4]. Nel documento vengono presentati gli obiettivi per lo sviluppo della rete europea di biblioteche digitali e per la creazione di uno spazio informativo comune, definendo un sistema regolamentare moderno per l'economia digitale che favorisse la disponibilità di contenuti digitali, l'investimento nella ricerca e nell'innovazione tecnologica per lo sviluppo di una società dell'informazione europea aperta e democratica. Il documento inoltre sottolineava l'importanza di un forte e concreto impegno da parte degli Stati membri.

In questo quadro, nel 2007, viene avviata Europeana, la Biblioteca digitale europea, il cui prototipo vede la luce nel 2008, mentre il servizio è pubblicato in rete nel 2010.

Si tratta di una grande e ambiziosa iniziativa gestita dalla Europeana Foundation, con il supporto della Commissione europea, che ha come obiettivo la raccolta e la diffusione di informazioni sul ricchissimo patrimonio culturale europeo tramite una piattaforma informatica con interfaccia multilingue.

Importantissimo il contributo di tutti gli Stati membri che favoriscono e supportano le istituzioni culturali, gli archivi, le biblioteche, i musei, gli archivi audiovisivi nella partecipazione a Europeana attraverso il conferimento dei metadati relativi alle risorse digitali detenute e gestite.

La sfida, allora, era quella di riempire di contenuti la Biblioteca digitale europea con una particolare attenzione alla qualità dei dati, alla definizione di standard condivisi per l'interoperabilità dei dati e alle regole comuni per l'accesso ai contenuti stessi.

La Commissione, quindi, agisce su due piani: da un lato intende rafforzare il quadro politico di riferimento attraverso l'emanazione di raccomandazioni[5], dall'altro avvia un vasto piano di progetti europei che finanziano prevalentemente l'invio a Europeana dei contenuti sotto forma di metadati, favorendo la partecipazione attiva delle istituzioni culturali europee.

---

[4] https://eur-lex.europa.eu/EN/legal-content/summary/i2010-digital-libraries.html.

[5] Raccomandazione dell'agosto 2006 della Commissione europea dal titolo: *Commission reccomendation on the digitisation and online accessibility of cultural material and digital preservation* (2006/585/EC), <http://data.europa.eu/eli/reco/2006/585/oj>; Raccomandazione del 27 ottobre 2011 sulla digitalizzazione e l'accessibilità in rete dei materiali culturali e sulla conservazione digitale emessa nel quadro dell'agenda digitale europea: *Commission reccomendation on the digitisation and online accessibility of cultural material and digital preservation (2011/711/EU),* <http://data.europa.eu/eli/reco/2011/711/oj>; Raccomandazione del novembre 2021 della CE (segue e rafforza quella del 2011) per lo sviluppo di uno spazio comune europeo dei dati nel settore della cultura: *Commission Recommendation on a common European data space for cultural heritage* (2021/1970/EU), <http://data.europa.eu/eli/reco/2021/1970/oj>. La raccomandazione rinnovava l'esortazione agli stati membri ad accelerare la protezione e la conservazione del patrimonio culturale sostenendo Europeana, un'infrastruttura digitale sicura e sostenibile.

MODUS

Sul piano dei progetti europei il Ministero italiano, attraverso l'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU), ha partecipato attivamente alla costruzione della Biblioteca digitale europea. Sulla base dei risultati positivi ottenuti dal progetto MINERVA venivano quindi avviati e realizzati altri progetti europei.
I principali progetti, coordinati dall'ICCU per conto del Ministero, sono stati Athena (2008-20011), Linked Heritage (2011-2013) e AthenaPlus (2013-2015), che vengono realizzati nell'arco di otto anni. Sono progetti di grande rilievo, sia per la larga partecipazione di istituti culturali appartenenti a un gran numero di paesi europei (20 paesi europei e due extraeuropei - la Russia e Israele - e centinaia di istituti culturali), sia per i risultati raggiunti. I tre progetti hanno inviato a Europeana oltre sei milioni di dati. Tra le attività svolte, va ricordata la produzione di linee guida e strumenti tecnologici ampiamente usati da Europeana e da altri progetti europei che a essa contribuiscono. Si tratta di standard terminologici, catalografici e di descrizione delle risorse digitali prodotte dai musei e dagli altri istituti culturali, che hanno fornito un contributo concreto al processo di costituzione di Europeana[6].
Inoltre, nell'ambito dei tre progetti e in collaborazione con Europeana, è stata avviata una discussione su importanti tematiche relative al patrimonio culturale digitale quali la gestione delle risorse digitali, la sostenibilità, la conservazione a lungo termine e il riuso di tali contenuti per l'istruzione, la ricerca, il tempo libero e il turismo. Tra i vari esempi di riuso, vi sono le mostre digitali. A tal proposito l'ICCU ha realizzato un kit open source gratuito per la progettazione di mostre virtuali fruibili via web o piattaforme mobili. Si tratta del progetto MOVIO[7] finanziato nell'ambito di un bando lanciato nel 2011 dalla Fondazione Telecom Italia e successivamente arricchito da AthenaPlus con l'elaborazione di linee guida sul tema. Con l'uso del kit MOVIO le istituzioni culturali sono in grado di realizzare mostre digitali riutilizzando e valorizzando le proprie risorse digitali o quelle recuperate in rete.
In relazione al riuso dei contenuti digitali si sviluppa un dibattito sul quadro giuridico di disciplina della licenza delle opere protette da copyright. Da qui infatti deriva la difficoltà di pubblicare in rete opere protette con una conseguente scarsità di opere del XX e del

[6] Sul sito <https://michael-culture.eu/library/> si possono trovare notizie sui progetti insieme ai testi scaricabili delle linee guida e dei numerosi documenti prodotti, tra i quali: *Metadata for the description of digital exhibitions*, <https://michael-culture.eu/mca-library/athena-plus-metadata-for-the-description-of-digital-exhibitions-the-demes-element-set/>; *Guidelines for the creation of digital exhibitions*, *Guidelines for geographical information*, <https://michael-culture.eu/mca-library/guidelines-for-geographic-information/>; *Digitization: standards landscape for European museums, libraries, archives*, <https://michael-culture.eu/mca-library/athena-digitisation-standards-landscape-for-european-museums-archives-libraries/>; *Persistent identifiers: recommendations*, <https://michael-culture.eu/mca-library/athena-persistent-identifiers-persistent-identifiers-recommendations/>; *Recommendations and guidelines for digital storytelling projects*, *Booklet for implementing LIDO*, <https://michael-culture.eu/mca-library/athena-plus-implementing-lido/>; *Lightweight Information Describing Object: LIDO*, <https://michael-culture.eu/mca-library/lightweight-information-describing-object-lido/>.
Inoltre sul sito di MINERVA, conservato da Internet Archive, <https://web.archive.org/web/20240507091135/https://www.minervaeurope.org/> si trovano notizie e documenti su Linked Heritage e AthenaPlus.

[7] www.movio.beniculturali.it.

XXI secolo in Europeana, che quindi è a favore di normative più flessibili e auspica che ai dati culturali siano applicate licenze aperte per consentire il loro riuso.
Dal settembre 2012 i metadati aggregati da Europeana sono rilasciati sotto licenza Creative Commons 0 (CC0), per renderli liberamente disponibili per il riutilizzo anche commerciale. Attraverso numerosi progetti europei aumentano i contenuti e cresce la rete delle istituzioni partecipanti a Europeana. Nasce quindi l'esigenza di aggregatori nazionali o tematici che supportino il lavoro delle istituzioni culturali, facilitando il conferimento dei dati alla Biblioteca digitale europea e il controllo di qualità dei dati stessi. Europeana viene infatti alimentata attraverso diversi canali, in primo luogo gli aggregatori nazionali, sviluppati in quasi tutti i paesi europei. Un aggregatore è un organismo che raccoglie metadati da una serie di fornitori di contenuti per renderli interoperabili con Europeana o con altre piattaforme, a seguito di procedure standard nel trattamento delle informazioni, nel formato dei file e nei diversi protocolli informatici. In genere, gli aggregatori svolgono anche un ruolo di supporto tecnico e formativo agli istituti culturali, che vengono seguiti durante tutto il processo. Questo ruolo è svolto in Italia da CulturaItalia[8], il portale della cultura gestito dall'ICCU.
Negli anni, infatti, l'ICCU sviluppa e consolida il ruolo di punto di riferimento nazionale per Europeana, sia attraverso la partecipazione ai citati progetti europei che a numerosi altri (per cui si rimanda al sito dell'Istituto nella sezione dedicata ai servizi), sia per la creazione e gestione di CulturaItalia, l'aggregatore nazionale trasversale che include risorse degli archivi, delle biblioteche, dei musei e di altre istituzioni culturali e di ricerca, ed è un'iniziativa condivisa con Regioni, Università ed altri importanti istituti culturali italiani.
Infine, vale la pena ricordare un altro progetto europeo coordinato dal Ministero della cultura, basato anche questo sugli standard prodotti da MINERVA. Si tratta del progetto Multilingual Inventiry of Cultural Heritage in Europe (MICHAEL) che ha avuto come obiettivo principale il censimento e la raccolta delle collezioni digitali a partire da inventari nazionali o regionali e lo sviluppo di una infrastruttura tecnologica basata sugli standard per la creazione, l'aggregazione e la pubblicazione dei dati, rendendo accessibile in rete contenuti digitali eterogenei, oltre a permettere il controllo sulle attività di digitalizzazione anche al fine di evitare duplicazioni e frammentazioni. Uno dei principali risultati è di aver creato nel 2007 un'associazione no-profit (una AISBL, Association International Sans But Lucratif) di diritto belga, Michael Cultural Association (MCA)[9]. I soci fondatori sono stati il Ministero italiano, il Ministero francese e il britannico Museums, Libraries and Archives Council (MLA), agenzia governativa per la digitalizzazione del patrimonio culturale. Oggi MCA è un network intersettoriale che si occupa di conservazione, promozione e valorizzazione dei contenuti culturali digitali. Attualmente conta più di 200 membri tra musei, biblioteche, archivi, centri di ricerca, ministeri dell'UE e anche oltre, e si avvale della collaborazione di numerosi esperti che operano nel settore del patrimonio culturale

[8] www.culturaitalia.it.
[9] www.michael-culture.eu.

digitale. Partecipa a molti progetti europei e recentemente ha ottenuto il supporto stabile della Commissione europea che ha riconosciuto il valore della rete creata e gestita dall'Associazione.

Per concludere, con uno sguardo al presente, è utile menzionare la strategia di Europeana 2020-2025[10]. La visione della Biblioteca digitale europea punta alla trasformazione digitale del settore del patrimonio culturale al fine di favorirne l'uso per l'istruzione, la ricerca, il tempo libero.

Le tre priorità individuate sono: rinforzare l'infrastruttura rendendola al passo con lo sviluppo tecnologico; migliorare la qualità dei dati utilizzando anche gli algoritmi per l'arricchimento dei metadati; sviluppare le competenze e incoraggiare l'adozione di standard condivisi, diffondendo le buone pratiche e stimolando soluzioni comuni. L'obiettivo è quindi consolidare i risultati raggiunti, sviluppare strumenti e politiche per accompagnare il cambiamento digitale, anche attraverso partnership private che favoriscano l'innovazione e contribuire ad una società della conoscenza aperta e creativa. Alla rivista DigItalia è affidato il compito di continuare a seguire il dibattito europeo, stimolando riflessioni e approfondimenti nel contesto italiano, analizzando le criticità e dando altresì visibilità alle tante iniziative, agli importanti sviluppi e ai risultati raggiunti a livello nazionale nel quadro delle politiche e delle strategie europee.

[10] https://pro.europeana.eu/page/strategy-2020-2025-summary.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

Una rivista per il dialogo interdisciplinare

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00139

Simonetta Buttò
Già direttrice dell'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU)

Nata alla fine del 2005 e pubblicata anche online in formato pdf su un sito dedicato, «DigItalia», a partire dal sottotitolo *Rivista del digitale nei beni culturali*, si collocava – nella visione del suo fondatore, Marco Paoli – al centro di un vasto territorio di discussione e di ricerca, ancora in parte inesplorato: da un lato la riflessione sull'uso delle tecnologie digitali per la conoscenza e la diffusione del patrimonio culturale, dall'altro l'adozione di un'ottica dichiaratamente transdisciplinare. Le due aspirazioni naturalmente convergevano, trattandosi del futuro dei beni culturali nel nostro Paese, in linea con quanto anche la Comunità europea auspicava in quegli anni, a cominciare dall'esigenza di coordinamento delle esperienze di digitalizzazione espressa nei Principi di Lund del 2001[1].

L'orizzonte del nuovo progetto editoriale dell'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU) era molto esteso e insieme molto specialistico, tanto da suscitare all'inizio qualche dubbio sulla sua sostenibilità[2] (smentito poi dai fatti), ma era soprattutto originale e all'avanguardia, in anni in cui il dialogo internazionale sul digitale per il patrimonio culturale era ancora all'inizio e il dibattito sull'integrazione dei saperi specialistici nei diversi settori dei beni culturali di là da venire: la politica culturale in Italia era in quegli anni (e lo sarebbe rimasta ancora molto a lungo) ancorata alla valorizzazione delle prerogative specifiche dei diversi domini (archivistico, storico artistico, archeologico, architettonico e librario) del settore, sia per quanto riguardava la prospettiva dei progetti nazionali, che per gli scambi e le relazioni in ambito europeo.

Anche se la base dei professionisti dei beni culturali, soprattutto bibliotecari e archivisti, aveva intuito già da tempo la vitalità e le opportunità offerte da forme di dialogo interdisciplinare, teorico e pratico, la fondazione ufficiale del coordinamento permanente MAB (Musei Archivi Biblioteche) voluto dall'Associazione Italiana Biblioteche (AIB), dall'Associazione Nazionale Archivistica Italiana (ANAI) e da ICOM Italia, vedrà la luce solo nel 2011, e nel corso degli anni stenterà non poco a individuare possibili modelli di mediazione fra competenze comuni e competenze specialistiche, in un contesto domi-

[1] I *Principi di Lund* del 4 aprile 2001 si possono leggere in traduzione italiana sul sito dell'ICCU: <https://www.iccu.sbn.it/export/sites/iccu/documenti/lund_principles-it.pdf>.
Sul recepimento del Piano di azione di Lund cfr. Rossella Caffo, *La carta di Parma per la diffusione della cultura europea sul web*, «Bollettino AIB», 43 (2003), n. 4, p. 489-496, <https://bollettino.aib.it/article/view/5116>.

[2] Philippe Marcerou, *Digitalia: rivista del digitale nei beni culturali, n. 0, 2005*, «Bulletin des Bibliothéques de France», 2006, n. 6, p. 106-107, <https://bbf.enssib.fr/consulter/bbf-2006-06-0106-005>.

MODUS

nato da solide specificità professionali e acquisite tradizioni disciplinari, nonostante la buona volontà dei singoli, lasciando indietro una riflessione più ampia, e comune, sulla funzione sociale della gestione del patrimonio culturale, nell'interesse degli utenti e delle comunità di cui essi sono parte.

A dieci anni dalla sua prima apparizione, tuttavia, «Digitalia», che grazie al supporto dell'Università di Macerata utilizzava dal 2012 la piattaforma OJS (Open Journal Systems) dotata di maggiori funzionalità di ricerca e navigazione, aveva all'attivo una crescente reputazione nel campo dell'editoria periodica specializzata, non solo in Italia, conservando come obiettivo primario quello di dare voce al dibattito sul tema dell'applicazione delle tecnologie digitali ai diversi settori del patrimonio culturale presentando progetti – nazionali e internazionali – in corso di svolgimento o conclusi, interrogandosi sull'esperienza degli utenti, sugli standard e le metodologie, sui problemi di sostenibilità nel lungo termine e fornendo documentazione e linee guida prodotte in ambito internazionale. Fondamentale, da questo punto di vista, è stata l'intensa partecipazione dell'ICCU, capofila degli istituti di cultura italiani del Ministero nell'ambito delle politiche europee di digitalizzazione del patrimonio culturale, a numerosi progetti europei, in qualità di coordinatore o collaboratore di estesi network di ricerca transnazionali.

Accreditata negli elenchi ANVUR (Agenzia nazionale di valutazione del sistema universitario e della ricerca) come "rivista scientifica" per le discipline dell'Area 11 (Scienze storiche, filosofiche, pedagogiche e psicologiche), «DigItalia» negli anni aveva pubblicato articoli sul digitale nei beni culturali per il settore storico, museale, archeologico, archivistico e – naturalmente – librario, offrendo una sede ideale di discussione a quanti, nel contesto internazionale e in Italia, all'interno del Ministero e al di fuori di esso, nella vasta platea degli istituti culturali e nelle università, portavano avanti progetti tecnologicamente avanzati di digitalizzazione del patrimonio e piattaforme innovative per l'accesso aperto a tutte le fasce di pubblico.

Grazie alla solidità dei servizi bibliografici nazionali gestiti dall'ICCU, e in particolare del catalogo del Servizio Bibliotecario Nazionale (SBN) – giunto al suo trentesimo anno di attività e in costante crescita –, del *Censimento dei manoscritti antichi e moderni* Manus Online (MOL), del *Censimento nazionale delle edizioni italiane del XVI secolo* (EDIT16) e della Biblioteca Digitale Italiana (BDI) del portale Internet Culturale, nel 2015 la comunità italiana delle biblioteche poteva vantare un'esperienza delle nuove tecnologie tale da consentire di pensare al futuro della lettura, della ricerca e della conoscenza del patrimonio culturale, allargando l'orizzonte alle contemporanee buone pratiche svolte in ambito internazionale.

A partire dal volume monografico del 2017 dedicato alla *Conferenza nazionale* per il trentennale di SBN[3] e ad analoghe celebrazioni sui territori e presso la comunità delle biblioteche ecclesiastiche[4], era naturale che «DigItalia» si prestasse a rappresentare anche

[3] *Conferenza Nazionale 1986-2016: 30 anni di biblioteche in rete: Roma, Biblioteca nazionale centrale, 1° aprile 2016: atti*, «DigItalia. Rivista del digitale nei beni culturali», 12 (2017), n. 1/2, p. 14-69.

[4] Ivi, p. 72-215.

MODUS

la sede ideale per il dibattito in corso, in concomitanza con l'avvio di una riconsiderazione dei servizi bibliografici nazionali, che prendeva vita in quel periodo, grazie ai risultati presentati da tre gruppi di lavoro costituiti dall'ICCU e dedicati rispettivamente all'*Evoluzione e sviluppo di SBN*, alle *Infrastrutture per il patrimonio bibliografico e digitale* e alle *Linee d'azione per la definizione delle politiche per l'accesso ai servizi*[5].

L'attenzione si rivolgeva dunque in questa prima fase al potenziamento del processo di integrazione tra dati bibliografici e contenuti digitali all'interno delle basi dati nazionali, nella convinzione che fosse necessario un superamento della frammentazione e della pluralità degli strumenti per l'accesso all'informazione e l'avvio di un processo di fruizione multimediale più estesa.

Nasceva da questa riflessione il nuovo progetto lanciato dall'ICCU nel 2019[6] che prevedeva la costruzione di un "Sistema di ricerca integrato" dotato di un punto di accesso unico, il portale Alphabetica[7], che consentiva la consultazione contemporanea di record e risorse digitali provenienti dalla grande comunità di SBN, ma anche – grazie all'apertura e alla flessibilità del sistema – al di fuori di essa. In breve tempo sono stati integrati nel sistema la base dati *14-18: documenti e immagini della Grande Guerra*, ricca di testimonianze a stampa e manoscritte, diari, fotografie, immagini di monumenti e sacrari, documenti sonori sulla prima guerra mondiale e il sito MOVIO (Mostre virtuali online) contenente mostre multimediali realizzate da archivi, scuole, istituzioni culturali e università, entrambe *cross domain* e gestite dall'ICCU. In convenzione con la Regione Emilia-Romagna sono state anche aggregate al portale le risorse grafiche, fotografiche e cartografiche dal XV al XXI secolo presenti nel catalogo collettivo IMAGO[8].

Se è vero che le biblioteche sono storicamente multimediali, conservando una pluralità di oggetti, disegni, fotografie, strumenti scientifici, monete e altri materiali non riconducibili al manufatto libro, il nuovo progetto coinvolgeva, e coinvolge, anche numerosi istituti culturali, associazioni, centri di ricerca presenti in Italia che non si identificano univocamente in un museo, una biblioteca, un archivio, ma spesso sono produttori di contenuti in una prospettiva "naturalmente" *cross domain*: un ulteriore passo in avanti nella fruizione diretta di contenuti culturali indipendentemente dalla visione settoriale del sistema delle biblioteche.

È fuori di dubbio che le tecnologie più avanzate, unite a una visione davvero allargata del patrimonio come risorsa pubblica, hanno offerto uno strumento formidabile per l'ampliamento del dialogo interdisciplinare, perché l'ambiente digitale per sua natura

[5] Per la composizione interistituzionale dei Gruppi di lavoro, nominati nel novembre del 2015, cfr. sul sito dell'ICCU: <https://www.iccu.sbn.it/it/attivita-servizi/gruppi-di-lavoro-e-commissioni/pagina_0007.html>; <https://www.iccu.sbn.it/it/attivita-servizi/gruppi-di-lavoro-e-commissioni/pagina_0008.html>; <https://www.iccu.sbn.it/it/attivita-servizi/gruppi-di-lavoro-e-commissioni/pagina_0009.html>.

[6] Cfr. *Il Portale delle biblioteche e degli istituti culturali italiani: presentazione del progetto, Roma, 11 aprile 2019, Sala Spadolini, MiBAC*, «DigItalia. Rivista del digitale nei beni culturali», 14 (2019), n. 1, p. 7-28.

[7] Per gli atti della giornata di presentazione del progetto cfr. *Alphabetica e il nuovo ecosistema dei servizi bibliografici nazionali: Ministero della cultura, via del Collegio romano 27, Sala Giovanni Spadolini, giovedì 16 dicembre 2021*, «DigItalia. Rivista del digitale nei beni culturali», 17 (2022), n. 1, p. 7-110.

[8] Cfr.: <https://imago.regione.emilia-romagna.it/opac/.do>.

MODUS

abilita la lettura plurima del patrimonio da parte del pubblico, lasciando che le informazioni necessarie alla descrizione del dato rimangano peculiari alla classe di appartenenza del materiale (sia bibliografico che archivistico o museale).
Negli anni più recenti questa esperienza di *Ecosistema digitale dei servizi bibliografici nazionali* realizzato dall'ICCU ha costituito la base solida, lo spunto di partenza, per il nuovo progetto lanciato dal Ministero nel 2020, con la nascita dell'Istituto centrale per la digitalizzazione del patrimonio culturale (Digital library) e la pubblicazione del Piano nazionale di digitalizzazione (PND)[9] in vista della costruzione di un *Ecosistema digitale nazionale*, *cross domain*, in grado di ospitare, valorizzare e gestire le risorse digitali di tutto il patrimonio culturale del nostro Paese in sicurezza.
Le tappe più importanti di questo percorso, tuttora *in itinere*, si possono seguire sulle pagine di «DigItalia», che ha dato puntualmente conto dello sviluppo ed evoluzione delle infrastrutture e dei servizi digitali per il patrimonio culturale, a disposizione delle istituzioni pubbliche e private per favorire l'interoperabilità tra i sistemi e far dialogare dati appartenenti a domini diversi della conoscenza, valorizzando il capitale semantico del patrimonio culturale che si fonda sull'intersezione dei saperi.
Grazie a un Comitato scientifico composto da esperti di alto livello, a una redazione colta, competente e appassionata e a ben oltre 500 autori, italiani e stranieri, appartenenti ai diversi ambiti disciplinari (archivi, biblioteche, musei) e al mondo dell'Università, la rivista che ha dato voce ai più importanti progetti europei e internazionali e alle nuove frontiere affrontate nel nostro Paese sul tema dell'accesso al patrimonio culturale digitale, portando un contributo decisivo all'affermazione del principio di titolarità espresso dalla Convenzione quadro del Consiglio d'Europa sul valore dell'eredità culturale per la società firmata a Faro nel 2005, per un approccio partecipativo e sostenibile alla tutela e alla valorizzazione del patrimonio.
A coronamento del suo percorso attraverso l'evoluzione del digitale nella riflessione teorico-pratica, «DigItalia» ha affrontato l'esame dell'ANVUR e nel 2022 è stata riconosciuta come periodico di Classe A per le aree 11/A1, Storia medievale, 11/A2, Storia moderna, 11/A4, Scienze del libro e del documento e scienze storico religiose e 10/A1, Archeologia. È stata inclusa inoltre negli elenchi delle pubblicazioni di carattere scientifico per tutte le aree 10, Scienze dell'antichità, filologico-letterarie e storico-artistiche oltre che, come si è detto, per quelle dell'area 11.
L'ambìto riconoscimento accademico, unito all'adozione di una politica per l'accesso aperto, rappresenta l'approdo di un'idea di successo, sostenuta costantemente nel tempo con determinazione e chiarezza di vedute, e la garanzia di un futuro di crescita nella divulgazione di contenuti di alto valore culturale e nel coinvolgimento attivo dell'intera comunità professionale.

[9] Cfr. <https://digitallibrary.cultura.gov.it/il-piano/>.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

Saggi

L'impatto dell'European Accessibility Act sul settore culturale: obblighi e opportunità per archivi, musei e biblioteche

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00140

Stefano Allegrezza
Università degli studi di Macerata

A partire dal 28 giugno 2025, l'European Accessibility Act (direttiva UE 2019/882) ha introdotto obblighi significativi in materia di accessibilità digitale anche per gli istituti culturali, pubblici e privati. Dopo un inquadramento normativo e concettuale sul tema dell'accessibilità digitale, l'articolo ricostruisce l'evoluzione legislativa europea e nazionale, soffermandosi sull'interazione tra la nuova direttiva e la precedente direttiva UE 2016/2102. Vengono poi analizzate le ricadute dirette e indirette per i soggetti culturali, distinguendo alcuni ambiti applicativi. Viene evidenziato come l'accessibilità non debba essere considerata un mero vincolo tecnico o normativo, bensì un'opportunità per rafforzare il ruolo democratico e inclusivo delle istituzioni culturali. In un contesto sempre più attento alla disabilità e all'inclusione sociale, la direttiva, pur presentando alcune criticità interpretative, offre un'occasione unica per ripensare l'accessibilità come paradigma progettuale e missione democratica. Il saggio si conclude sottolineando l'urgenza di un adeguamento culturale, volto a integrare pienamente l'accessibilità nelle strategie digitali e inclusive dei luoghi della cultura.

1. Introduzione

La data del 28 giugno 2025 ha segnato un punto di svolta non solo per gli esperti di accessibilità digitale, ma anche per i professionisti che operano negli archivi, nelle biblioteche e nei musei. A partire da quella data, infatti, è entrata ufficialmente in vigore la direttiva UE 2019/882[1], meglio nota come *European Accessibility Act* (EAA) – recepita in Italia dal decreto legislativo 27 maggio 2022, n. 82 – che stabilisce i requisiti di accessibilità per una serie di prodotti e servizi, contribuendo a rimuovere barriere ancora molto diffuse nel mondo digitale e introducendo importanti novità anche per il settore culturale. In un contesto in cui sempre più servizi, documenti e interazioni avvengono in forma digitale, garantire l'accessibilità universale è cruciale per evitare nuove forme di esclusione sociale.

[1] Cfr. *Direttiva UE 2019/882 del Parlamento europeo e del Consiglio del 17 aprile 2019 sui requisiti di accessibilità dei prodotti e dei servizi*, <https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32019L0882>.

MODUS

Questo articolo intende analizzare le implicazioni dell'European Accessibility Act per le istituzioni culturali e i luoghi della cultura, come i musei, le biblioteche, gli archivi, le aree e i parchi archeologici, i complessi monumentali[2], che offrono servizi digitali, ad esempio tramite siti web e altre piattaforme online.

2. Cosa si intende per accessibilità digitale?

Per "accessibilità (digitale)" si intende «la capacità dei sistemi informatici ivi inclusi i siti web e le applicazioni mobili, nelle forme e nei limiti consentiti dalle conoscenze tecnologiche, di erogare servizi e fornire informazioni fruibili, senza discriminazioni, anche da parte di coloro che a causa di disabilità necessitano di tecnologie assistive o configurazioni particolari»[3]. In altre parole, l'accessibilità digitale assicura a qualunque persona, indipendentemente dalle sue condizioni e dal suo stato di salute, di poter utilizzare gli strumenti digitali senza incontrare difficoltà. Per fare un esempio, se un sito web è accessibile, una persona cieca o ipovedente può navigarlo per mezzo di una tecnologia assistiva, come uno screen reader, ovvero un software che legge ad alta voce tutto ciò che compare sullo schermo, dagli elementi della navigazione ai testi, ai titoli e perfino alle descrizioni delle immagini[4]; una persona con disturbi della sfera

[2] Cfr. decreto legislativo 22 gennaio 2004, n. 42, *Codice dei beni culturali e del paesaggio, ai sensi dell'articolo 10 della legge 6 luglio 2002, n. 137*, art. 101, *Istituti e luoghi della cultura*. «1. Ai fini del presente codice sono istituti e luoghi della cultura i musei, le biblioteche e gli archivi, le aree e i parchi archeologici, i complessi monumentali. 2. Si intende per: a) "museo", una struttura permanente che acquisisce, cataloga, conserva, ordina ed espone beni culturali per finalità di educazione e di studio; b) "biblioteca", una struttura permanente che raccoglie, cataloga e conserva un insieme organizzato di libri, materiali e informazioni, comunque editi o pubblicati su qualunque supporto, e ne assicura la consultazione al fine di promuovere la lettura e lo studio; c) "archivio", una struttura permanente che raccoglie, inventaria e conserva documenti originali di interesse storico e ne assicura la consultazione per finalità di studio e di ricerca; d) "area archeologica", un sito caratterizzato dalla presenza di resti di natura fossile o di manufatti o strutture preistorici o di età antica; e) "parco archeologico", un ambito territoriale caratterizzato da importanti evidenze archeologiche e dalla compresenza di valori storici, paesaggistici o ambientali, attrezzato come museo all'aperto; f) "complesso monumentale", un insieme formato da una pluralità di fabbricati edificati anche in epoche diverse, che con il tempo hanno acquisito, come insieme, una autonoma rilevanza artistica, storica o etnoantropologica».

[3] Cfr. art. 2 della legge 9 gennaio 2004, n. 4, *Disposizioni per favorire e semplificare l'accesso degli utenti e, in particolare, delle persone con disabilità agli strumenti informatici*.

[4] Uno *screen reader* (in italiano: lettore di schermo) è un software che consente alle persone con disabilità visive — cieche o ipovedenti — di accedere ai contenuti presenti su uno schermo, come quello di un computer, uno smartphone o un tablet. Lo screen reader interpreta il contenuto testuale e, tramite sintesi vocale (output audio) o display braille (output tattile), lo rende fruibile all'utente. Di solito non si limita a leggere semplicemente il testo visibile, ma fornisce anche informazioni sulla struttura della pagina (come titoli e intestazioni, descrizioni alternative delle immagini, paragrafi e link, tabelle e liste, moduli e campi da compilare, elementi di navigazione come i menu, elementi attivi come pulsanti, radio button ecc.) in modo da facilitare l'interazione con l'utente. A tal fine, lo screen reader sfrutta le informazioni semantiche fornite dal codice HTML o dai metadati dei documenti. Quando questi dati non sono presenti (come avviene, ad esempio, nei PDF non taggati, o nelle immagini prive di descrizione ecc.), il contenuto può risultare inaccessibile. Alcuni esempi di screen reader noti sono JAWS (Job Access With Speech), NVDA (NonVisual Desktop Access), VoiceOver (integrato in tutti i dispositivi Apple: macOS, iOS, iPadOS), TalkBack (integrato nei dispositivi Android), Orca (screen reader open source per Linux). Gli screen reader sono strumenti fondamentali per garantire il diritto all'informazione e all'uso autonomo delle tecnologie da parte delle persone cieche. Per questo, le normative sull'accessibilità digitale (come la direttiva UE 2016/2102 o la direttiva UE 2019/882) impongono che siti web, app e documenti digitali siano compatibili con questi strumenti.

MODUS

visiva può ingrandire le parti che gli interessano in modo da leggerle correttamente; una persona daltonica può comprendere tutti i particolari grazie a una scelta accurata dei colori; una persona con dislessia o che ha poca dimestichezza con la lingua può comprendere i contenuti e ottenere tutte le informazioni che cerca.
Inoltre, un sito accessibile permette l'interazione anche alle persone che per qualche motivo non possono usare il mouse, essendo navigabile anche solo con il tabulatore della tastiera.
Le persone sorde possono incontrare problemi a sentire i contenuti multimediali di un sito, per esempio un video o un webinar che non preveda i sottotitoli o un podcast che non riporti il testo del contributo audio: in un sito accessibile tutti i documenti audiovisivi avranno i sottotitoli per chi non sente e, viceversa, le audio descrizioni per chi non vede.
Oltre ai siti web, l'accessibilità digitale si riferisce anche alle applicazioni per dispositivi mobili (app) che si usano sullo smartphone, uno strumento che ormai fa parte della vita delle persone e riveste una grande utilità per tutti, ma specialmente per coloro che hanno qualche difficoltà e potrebbero facilmente accedere a una serie di servizi senza muoversi da casa.
Si tratta di una platea di utenti che non è affatto trascurabile: secondo l'Organizzazione mondiale della sanità (OMS) in tutto il mondo oltre 1.3 miliardi di persone soffrono di disabilità significative[5]. Ciò rappresenta il 16% della popolazione mondiale, ovvero una persona su sei. Per quanto riguarda l'Europa, secondo le stime di Eurostat, il 27% della popolazione sopra i sedici anni vive con una qualche forma di disabilità[6] e ciò equivale a circa centouno milioni di persone, ovvero un adulto su quattro nell'Unione Europea. In Italia, le persone con disabilità sono quasi tredici milioni, di cui oltre tre milioni in condizioni di grave disabilità. Complessivamente, si parla di circa il 22% della popolazione italiana, poco più di un italiano su cinque. Naturalmente, una buona parte delle persone che possono incontrare difficoltà nella navigazione e lettura di un sito o di un'app non accessibile sono quelle anziane, che possono avere problemi di vista, di udito, di abilità manuale[7]. In sostanza, nel terzo millennio circa il 20% delle persone non riesce ad accedere pienamente ai contenuti digitali a causa di qualche difficoltà sensoriale, di movimento, di comprensione o culturale. Secondo le previsioni, il numero delle persone con disabilità, anche temporanea, dovrebbe aumentare in modo significativo nell'Unione europea, anche in virtù dell'allungamento della vita e del progressivo invecchiamento della popolazione.

[5] Organizzazione mondiale della sanità (OMS). *Disability and Health fact sheet*, 7 marzo 2023, <https://www.who.int/health-topics/disability>.
[6] Cfr. Eurostat, *Population with disability*, luglio 2024, <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_with_disability>.
[7] Anche per questo motivo i portali di e-commerce più famosi come Amazon e i social network più diffusi hanno già provveduto a rendere accessibili i propri siti, non solo per una motivazione etica ma anche per poter allargare i propri utenti a una maggiore quota di pubblico.

3. Il contesto normativo

Il fondamento normativo internazionale per l'accessibilità, compresa quella digitale, è costituito dalla *Convenzione delle Nazioni Unite sui diritti delle persone con disabilità*[8], approvata dall'Assemblea delle Nazioni unite nel dicembre 2006. Tale convenzione è stata ratificata dal Parlamento della Repubblica Italiana diventando legge dello Stato[9]. Il 23 dicembre 2010 la Convenzione è stata ratificata anche dall'Unione europea. Questo significa che l'inclusione delle persone con disabilità è un obiettivo che non può più essere trascurato ed è fondamentale per raggiungere gli Obiettivi di Sviluppo Sostenibile (*Sustainable Development Goals*)[10]. L'articolo 9 di tale Convenzione stabilisce che gli Stati firmatari devono garantire che le persone con disabilità abbiano accesso, su base di uguaglianza, all'ambiente fisico, ai trasporti, alle tecnologie dell'informazione e della comunicazione (ICT) e ad altri servizi aperti o forniti al pubblico. La Convenzione ha avuto un impatto significativo sulle politiche globali, fungendo da catalizzatore per l'adozione di legislazioni sovranazionali e nazionali. L'Unione europea, in quanto organizzazione sovranazionale parte della Convenzione, ha utilizzato questo quadro come punto di riferimento per sviluppare un corpus normativo incentrato sull'accessibilità digitale, adottando un approccio graduale.

Un primo passaggio cruciale è stata l'emanazione della direttiva UE 2016/2102[11], comunemente conosciuta come *Web Accessibility Directive* (WAD) che ha imposto agli enti pubblici degli Stati membri di rendere accessibili i propri siti web e le proprie applicazioni

[8] La Convenzione delle Nazioni Unite sui diritti delle persone con disabilità (in inglese: *Convention on the Rights of Persons with Disabilities*, CRPD) è un trattato internazionale adottato dalla 61-esima Assemblea Generale delle Nazioni Unite il 13 dicembre 2006 (Risoluzione n. A/RES/61/106) ed entrato in vigore il 3 maggio 2008. L'obiettivo della Convenzione è promuovere, proteggere e garantire il pieno ed eguale godimento di tutti i diritti umani e delle libertà fondamentali da parte delle persone con disabilità, e promuovere il rispetto della loro dignità intrinseca. La Convenzione si basa su una serie di principi fondamentali: Rispetto per la dignità intrinseca e per l'autonomia individuale; Non discriminazione; Piena ed effettiva partecipazione e inclusione nella società; Rispetto per la differenza e l'accettazione delle persone con disabilità come parte della diversità umana; Pari opportunità; Accessibilità; Parità tra uomini e donne; Rispetto per lo sviluppo delle capacità dei bambini con disabilità. Cfr. United Nations, *Convention on the Rights of Persons with Disabilities* (CRPD), <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities>.

[9] Cfr. legge 3 marzo 2009, n. 18, *Ratifica ed esecuzione della Convenzione delle Nazioni Unite sui diritti delle persone con disabilità, con Protocollo opzionale, fatta a New York il 13 dicembre 2006 e istituzione dell'Osservatorio nazionale sulla condizione delle persone con disabilità*.

[10] Cfr. l'Obiettivo di Sviluppo Sostenibile (SDG) n. 10 "Ridurre le disuguaglianze" dell'Agenda 2030 delle Nazioni Unite. In particolare, l'accessibilità è parte integrante del Target 10.2, che recita: «Entro il 2030, potenziare e promuovere l'inclusione sociale, economica e politica di tutte le persone, indipendentemente da età, sesso, disabilità, razza, etnia, origine, religione o status economico o altro», <https://asvis.it/goal-e-target-obiettivi-e-traguardi-per-il-2030>.

[11] Cfr. *Direttiva UE 2016/2102 del Parlamento europeo e del Consiglio, del 26 ottobre 2016, relativa all'accessibilità dei siti web e delle applicazioni mobili degli enti pubblici*, <https://eur-lex.europa.eu/legal-content/IT/ALL/?uri=CELEX:32016L2102>. La direttiva stabiliva scadenze diverse per l'adeguamento: i siti web pubblicati dopo il 22 settembre 2018 dovevano essere adeguati entro il 23 settembre 2019; per i siti web esistenti prima del 23 settembre 2018, invece, la direttiva concedeva più tempo per l'adeguamento, ossia fino al 23 settembre 2020; infine, per le applicazioni mobili, la scadenza per l'adeguamento era fissata al 23 giugno 2021. In sostanza, ad oggi, tutti i siti web (compresi i documenti pubblicati nelle varie sezioni, incluso l'albo online) e le applicazioni mobili degli enti pubblici dovrebbero essere accessibili.

MODUS

per dispositivi mobili conformemente agli standard WCAG 2.1 di livello AA[12]. La direttiva, entrata in vigore dal 22 dicembre 2016, doveva diventare legge negli Stati membri entro il 23 settembre 2018 ed è stata recepita in Italia con il decreto legislativo 10 agosto 2018, n. 106[13]. Tuttavia, tale direttiva si concentrava esclusivamente sul settore pubblico, lasciando il settore privato al di fuori del suo campo di applicazione. Ciò ha portato a un panorama normativo frammentato, con alcune aziende private che hanno adottato l'accessibilità su base volontaria mentre altre hanno ignorato del tutto la questione[14]. È per questi motivi che tre anni più tardi, con la direttiva UE 2019/882[15], l'Unione europea ha ampliato il campo di applicazione della normativa sull'accessibilità estendendola anche ai soggetti privati. A differenza della direttiva UE 2016/2102, che riguardava i siti web e le applicazioni per dispositivi mobili, la nuova direttiva, comunemente nota come European Accessibility Act (EAA), si applica a una gamma ancor più vasta di prodotti e servizi, tra cui le comunicazioni elettroniche, i terminali self-service, i servizi bancari e i documenti digitali elettronici associati a tali servizi. Tale direttiva è stata recepita in Italia con il decreto legislativo 27 maggio 2022, n. 82[16] che è entrato in vigore il 16 luglio 2022, ma le relative disposizioni sono efficaci a far data dal 28 giugno 2025.

---

[12] Cfr. World Wide Web Consortium (W3C), *Web Content Accessibility Guidelines (WCAG) 2.1*, 2018, <https://www.w3.org/TR/WCAG21>. Le WCAG prevedono tre livelli di conformità: il livello A (minimo) che prevede il soddisfacimento dei requisiti di base di accessibilità; il livello AA (intermedio), quello più comunemente richiesto per siti istituzionali e pubblici; il livello AAA (massimo), spesso difficile da raggiungere completamente. Un contenuto digitale conforme al livello AA deve soddisfare tutti i criteri stabiliti dal livello A e da quello AA, come un contrasto sufficiente tra testo e sfondo (minimo 4.5:1), la presenza di testi alternativi per le immagini significative, la navigazione accessibile da tastiera, la presenza di etichette e istruzioni comprensibili nei moduli, nessuna dipendenza esclusiva dal colore per trasmettere informazioni, la compatibilità con gli screen reader e altri strumenti assistivi. Oltre alle *Web Content Accessibility Guidelines* (WCAG 2.1), che forniscono le linee guida principali per garantire l'accessibilità di contenuti web e documenti digitali, l'implementazione dell'European Accessibility Act è strettamente collegata a una serie di standard tecnici e di normative complementari che è importante tenere presenti. Tra questi, lo standard europeo EN 301549, *Accessibility requirements for public procurement of ICT products and services in Europe*, che contiene una serie di riferimenti per garantire l'acquisto di prodotti hardware e software e di servizi digitali accessibili ed è stata recepita in Italia con la norma UNI EN 301549, *Requisiti di accessibilità per prodotti e servizi ICT* il 17 dicembre 2020.

[13] Cfr. decreto legislativo 10 agosto 2018, n. 106 *Attuazione della direttiva UE 2016/2102 relativa all'accessibilità dei siti web e delle applicazioni mobili degli enti pubblici*. Il titolo originale del decreto, "Riforma dell'attuazione della direttiva UE 2016/2102 relativa all'accessibilità dei siti web e delle applicazioni mobili degli enti pubblici", è stato modificato nella forma attualmente vigente con il Comunicato della Presidenza del Consiglio dei ministri, relativo al decreto legislativo 10 agosto 2018, n. 106, recante: «Riforma dell'attuazione della direttiva UE 2016/2102 relativa all'accessibilità dei siti web e delle applicazioni mobili degli enti pubblici», pubblicato nella Gazzetta Ufficiale n. 215 del 15 settembre 2018. Questo decreto ha integrato la legge n.4 del 9 gennaio 2004.

[14] Occorre riconoscere che anche a livello di pubbliche amministrazioni ha regnato una grande confusione, e molte di esse (per non dire la maggior parte) ha continuato a non rispettare pienamente i dettami sull'accessibilità previsti dalla direttiva e dal decreto legislativo di recepimento. Ad esempio, molte pubbliche amministrazioni hanno continuato a pubblicare regolarmente sui propri siti istituzionali, compreso l'albo online, documenti elettronici in formati non accessibili, come documenti in formato PDF – o persino in formato PDF/A – ottenuti dalla semplice scansione di originali cartacei, senza alcuna operazione di riconoscimento ottico dei caratteri (OCR), rendendoli di fatto completamente inaccessibili agli screen reader utilizzati dalle persone con disabilità visive.

[15] Cfr. *Direttiva UE 2019/882 del Parlamento europeo e del Consiglio del 17 aprile 2019 sui requisiti di accessibilità dei prodotti e dei servizi*, <https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32019L0882>.

[16] Cfr. decreto legislativo 27 maggio 2022, n. 82, *Attuazione della direttiva UE 2019/882 del Parlamento europeo e del Consiglio, del 17 aprile 2019, sui requisiti di accessibilità dei prodotti e dei servizi*.

L'European Accessibility Act si colloca nel contesto della *Strategia per i diritti delle persone con disabilità 2021-2030*, elaborata dalla Commissione europea per ampliare gli obiettivi definiti nella precedente Strategia 2010-2020 allo scopo di migliorare la qualità della vita delle persone con disabilità, costruendo un ambiente inclusivo e libero da barriere o discriminazioni. L'obiettivo è duplice: da una parte, quello di ridurre i fattori che alimentano l'esclusione sociale e la povertà garantendo parità di accesso, partecipazione e opportunità, in linea con i principi del *Pilastro europeo dei diritti sociali*[17], e promuovendo l'inclusione sociale delle persone con disabilità; dall'altra, creare un mercato unico per prodotti e servizi accessibili, eliminando barriere normative tra gli Stati membri. La scadenza fissata per il 28 giugno 2025 ha rappresentato un punto di svolta, dato che da quel momento in poi i prodotti e i servizi non conformi potrebbero essere esclusi dal mercato.

Per quanto riguarda l'Italia, non si può non ricordare che per tutelare le persone fragili o con disabilità, già nel 2004 – due anni prima della *Convenzione delle Nazioni Unite sui diritti delle persone con disabilità* – era stata emanata la legge 9 gennaio 2004, n. 4, *Disposizioni per favorire e semplificare l'accesso degli utenti e, in particolare, delle persone con disabilità agli strumenti informatici*, la quale mirava a garantire l'uguaglianza di accesso a tutti i servizi online, promuovendo l'inclusione delle persone con disabilità nel mondo digitale e assicurando che nessuno rimanesse escluso (cfr. Figura 1). La legge "Stanca" si basava sul principio fondamentale dell'uguaglianza di accesso sancito dall'articolo 3 della Costituzione italiana[18] ed è stata anche il riferimento normativo da cui poi sono nate le *Linee guida sull'accessibilità degli strumenti ICT* dell'Agenzia per l'Italia Digitale (AgID)[19]. Tale legge, che in precedenza trovava applicazione solo nei confronti di soggetti pubblici[20], si applica ora anche ai soggetti che offrono servizi al

[17] Cfr. Commissione europea, *Pilastro europeo dei diritti sociali*, dicembre 2017, <https://commission.europa.eu/system/files/2017-12/social-summit-european-pillar-social-rights-booklet_it.pdf>.

[18] È la così detta legge "Stanca" in quanto venne proposta da Lucio Stanca, allora Ministro per l'innovazione e le tecnologie. Tale legge ha delegato la definizione di regole e requisiti ad una serie di decreti attuativi, come il decreto del Presidente della Repubblica, 1° marzo 2005, n. 75, intitolato *Regolamento di attuazione della legge 9 gennaio 2004, n. 4 per favorire l'accesso dei soggetti disabili agli strumenti informatici*, che ha definito alcune tematiche, tra cui la possibilità, prevista dalla legge 4/2004, del rilascio di un bollino di conformità; il decreto ministeriale 8 luglio 2005, intitolato *Requisiti tecnici e i diversi livelli per l'accessibilità agli strumenti informatici*, che contiene una serie di allegati con i requisiti tecnici, le metodologie e i programmi di valutazione assistita per la verifica dell'accessibilità dei siti web; e il decreto ministeriale 30 aprile 2008 *Regole tecniche disciplinanti l'accessibilità agli strumenti didattici e formativi a favore degli alunni disabili*. La legge è stata poi aggiornata con il decreto ministeriale del 20 marzo 2013 che aggiorna l'allegato A relativo ai criteri e ai metodi per la verifica tecnica e i requisiti tecnici di accessibilità. Considerata la natura mutevole e in continuo divenire del tema dell'accessibilità, nel tempo la legge Stanca è stata aggiornata e continuerà sicuramente ad esserlo; la novità più importante è stata senz'altro la ricezione della direttiva UE 2016/2102, relativa all'accessibilità dei siti web e delle applicazioni mobili degli enti pubblici, e della direttiva UE 2019/882 che estende le regole sull'accessibilità anche al mondo privato e stabilisce l'obbligo di rispettare determinati requisiti di accessibilità per alcuni prodotti e servizi messi in commercio dal 28 giugno 2025.

[19] Cfr. Agenzia per l'Italia Digitale, *Linee guida accessibilità – PA*, <https://www.agid.gov.it/it/design-servizi/accessibilita/linee-guida-accessibilita-pa>. L'AgiD mantiene una sezione del proprio sito dedicata a questo tema: <https://www.agid.gov.it/it/ambiti-intervento/accessibilita-usabilita>.

[20] Ovvero: le Pubbliche amministrazioni (come specificato dall'art. 1 comma 2 del decreto legislativo 30 marzo 2001, n. 165); gli Enti pubblici economici; le Aziende private concessionarie di servizi pubblici; le Aziende municipalizzate regionali; gli Enti di assistenza e di riabilitazione pubblici; le Aziende di trasporto e di telecomunicazione a prevalente partecipazione di capitale pubblico; le Aziende appaltatrici di servizi informatici.

MODUS

pubblico, attraverso siti web o applicazioni per dispositivi mobili, con un fatturato medio, negli ultimi tre anni di attività, superiore a cinquecento milioni di euro[21].
L'Italia ha recepito l'European Accessibility Act con il decreto legislativo 27 maggio 2022, n. 82, che è entrato in vigore il 28 giugno 2025. Tuttavia, i fornitori di servizi rientranti nell'ambito di applicazione del decreto potranno continuare a erogare i propri servizi secondo le modalità precedentemente adottate fino al 28 giugno 2030, a condizione che utilizzino prodotti già in uso prima della data di entrata in vigore. Qualora invece venga introdotto un nuovo prodotto per l'erogazione dei servizi, il rispetto dei requisiti previsti dal decreto diventerà immediatamente obbligatorio.

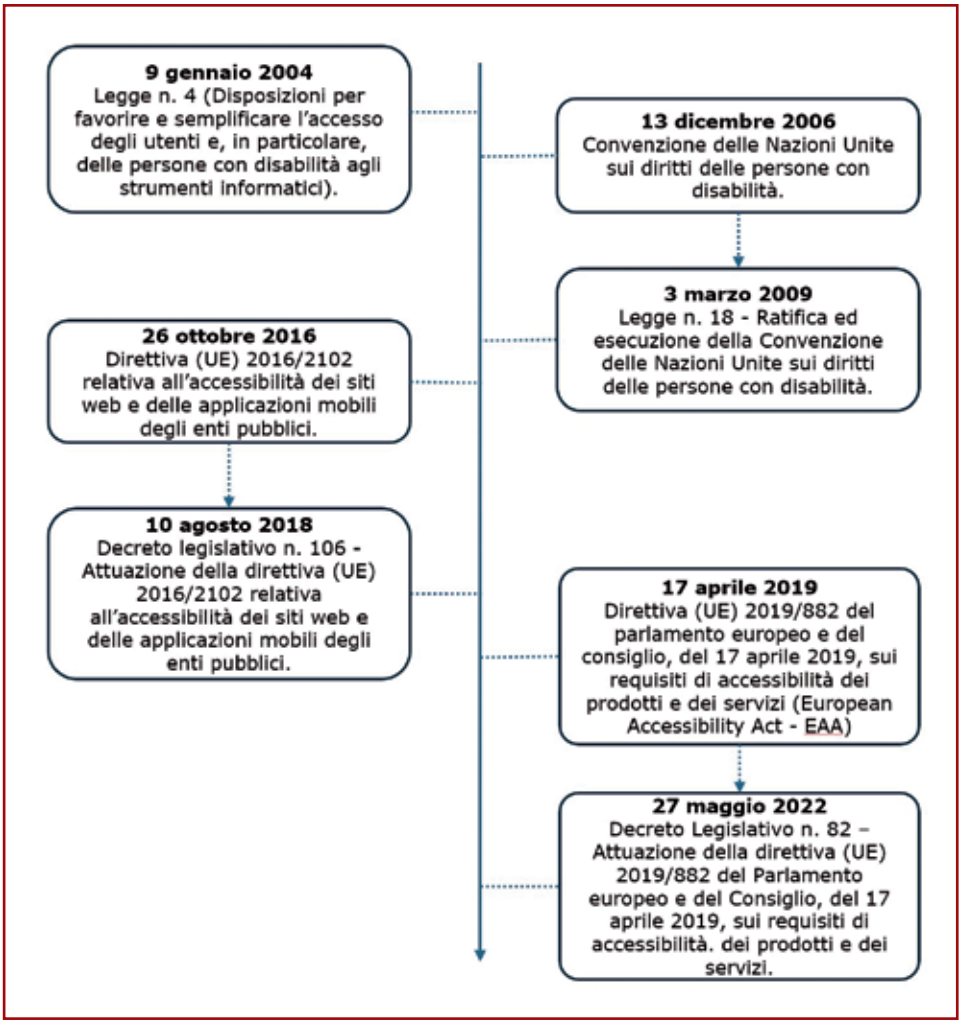

Figura 1. *Quadro normativo sintetico sull'accessibilità*

[21] In virtù del decreto legge 6 luglio 2020, n. 76, recante *Misure urgenti per la semplificazione e l'innovazione digitale* (c.d. decreto "Semplificazioni"), convertito con modificazioni dalla legge 11 settembre 2020, n. 120.

4. Categorie di prodotti e servizi coinvolti nell'European Accessibility Act

L'European Accessibility Act riguarda i prodotti immessi sul mercato e i servizi forniti a partire dal 28 giugno 2025. Per quanto riguarda la prima categoria, si applica[22] ai sistemi hardware e ai sistemi operativi informatici generici per consumatori; ai terminali self-service di pagamento e a quelli destinati alla fornitura di servizi quali sportelli automatici; alle macchine per l'emissione di biglietti; ai terminali per il check-in; ai terminali self-service interattivi destinati alla fornitura di informazioni; alle apparecchiature terminali con capacità informatiche interattive per consumatori utilizzate per i servizi di comunicazione elettronica; alle apparecchiature terminali con capacità informatiche interattive per consumatori utilizzate per accedere a servizi di media audiovisivi; ai lettori di libri elettronici (e-reader). Per quanto riguarda la seconda categoria, l'European Accessibility Act si applica[23] a una vasta gamma di servizi: servizi di comunicazione elettronica; servizi che forniscono accesso a servizi di media audiovisivi[24]; alcuni elementi relativi ai servizi di trasporto passeggeri (siti web; servizi per dispositivi mobili, comprese le applicazioni; biglietti elettronici e servizi di biglietteria elettronica; fornitura di informazioni relative ai servizi di trasporto, comprese le informazioni di viaggio in tempo reale, terminali self-service interattivi); servizi bancari; libri elettronici (e-book) e software dedicati[25]; servizi di commercio elettronico. Vi sono delle esclusioni, relative, ad esempio, ai media basati sul tempo[26] preregistrati e pubblicati prima del 28 giugno 2025; ai documenti in formati di file per ufficio (ad esempio, documenti creati con Microsoft Word, Excel o Powerpoint) pubblicati anch'essi prima del 28 giugno 2025; ai contenuti di siti web e applicazioni per dispositivi mobili considerati "archivi", nel senso che contengono soltanto contenuti che non sono stati aggiornati o rielaborati dopo il 28 giugno 2025.

---

[22] Per maggiori dettagli si rimanda all'art. 1, "Ambito di applicazione" del decreto legislativo 27 maggio 2022, n. 82.

[23] *Ibidem*.

[24] Per «servizi che forniscono accesso a servizi di media audiovisivi» si intendono i «servizi trasmessi da reti di comunicazione elettronica che sono utilizzati per individuare, selezionare, ricevere informazioni sui servizi di media audiovisivi e visualizzare tali servizi e tutte le caratteristiche correlate, quali sottotitoli per non udenti e ipoudenti, audiodescrizione, sottotitoli parlati e interpretazione in lingua dei segni, derivanti dall'attuazione di misure per rendere i servizi accessibili ai sensi dell'articolo 7 della direttiva 2010/13/UE; e includono guide elettroniche ai programmi (*electronic programme guides* — EPG)». Cfr. art. 3 "Definizioni" del decreto legislativo 27 maggio 2022, n. 82.

[25] A questo proposito va ricordata la direttiva UE 2017/1564 del Parlamento europeo e del Consiglio del 13 settembre 2017, che mira a garantire che le persone non vedenti, con disabilità visive o con altre difficoltà nella lettura di testi a stampa abbiano accesso ai libri e ad altri tipi di pubblicazioni – compresi gli spartiti musicali – ancorché protetti da diritti d'autore e diritti connessi. Tale direttiva è stata recepita in Italia con l'articolo 15 della L. n. 37/2019.

[26] L'espressione "media basati sul tempo" (*time-based media*, nella versione inglese) si riferisce a contenuti multimediali che si sviluppano nel tempo e che non possono essere pienamente compresi o fruiti senza seguire una sequenza temporale, a differenza, ad esempio, di un'immagine statica o di un documento testuale. Si tratta dei documenti sonori (audio), dei documenti audiovisivi (video), delle animazioni sincronizzate (inclusi filmati, podcast, webinar registrati ecc.). I contenuti multimediali preregistrati (non live) pubblicati prima del 28 giugno 2025 non sottostanno all'obbligo di dover essere resi accessibili secondo i criteri imposti dalla direttiva. Dopo quella data, invece, i nuovi contenuti dovranno rispettare i requisiti di accessibilità.

I requisiti specifici di accessibilità che i prodotti e servizi devono soddisfare sono dettagliati nell'Allegato I del decreto legislativo 27 maggio 2022, n. 82, (sezioni I-II per i prodotti, III-IV per i servizi), in linea con quanto previsto dalla direttiva. L'Allegato II presenta alcuni esempi, non vincolanti, di possibili soluzioni che contribuiscono a soddisfare i requisiti di accessibilità. In generale, tali requisiti riguardano aspetti progettuali volti a garantire che il prodotto o servizio sia fruibile anche da persone con diverse disabilità. Tra questi si possono citare: la disponibilità di alternative testuali per i contenuti non testuali (per esempio, le immagini), la fornitura di informazioni e istruzioni tramite più canali sensoriali, l'uso di interfacce compatibili con le tecnologie assistive, nonché la realizzazione di imballaggi e documentazione d'uso accessibili.
Va detto che la direttiva si applica solo ai prodotti sul mercato e ai servizi forniti dopo il 28 giugno 2025; i fornitori di servizi «possono continuare a prestare i loro servizi utilizzando prodotti che utilizzavano in modo legittimo prima di tale data per fornire servizi analoghi»[27] fino al 28 giugno 2030, potendo, quindi, godere di un tempo maggiore per l'adeguamento.

5. Soggetti nei cui confronti si applica l'European Accessibility Act

L'European Accessibility Act si applica direttamente alle specifiche categorie di prodotti e servizi descritte nel paragrafo precedente, fornite da persone fisiche e giuridiche, senza distinzione sulla natura pubblica o privata del fornitore[28]. Essendo l'ambito pubblico già disciplinato dalla direttiva UE 2016/2102, l'European Accessibility Act rafforza e amplia l'ambito, includendo anche i prodotti e i servizi offerti da soggetti privati. Gli istituti e i luoghi della cultura, sia pubblici che privati, rientrano pienamente nel campo di applicazione dell'European Accessibility Act, pur non essendo esplicitamente indicati come destinatari principali.
Tuttavia, per quanto riguarda gli obblighi di accessibilità, occorre fare una distinzione. Nel caso in cui l'istituto culturale rientri nella categoria della pubblica amministrazione (ad esempio, archivi di Stato, biblioteche pubbliche, musei civici ecc.) la direttiva UE 2016/2102 imponeva già l'obbligo di rendere accessibili i siti web e le applicazioni per dispositivi mobili, mentre l'European Accessibility Act impone l'obbligo di accessibilità per prodotti e servizi che rientrano nelle categorie viste precedentemente. Al contrario, nel caso in cui l'istituto culturale sia un soggetto privato (come fondazioni, enti del terzo settore, istituti ecclesiastici), non sussiste un obbligo generalizzato di rendere accessibili i siti web e le applicazioni mobili e l'European Accessibility Act impone l'obbligo di rendere accessibili solo gli specifici prodotti e servizi digitali ricompresi nel campo di applicazione della direttiva (ad esempio, siti di e-commerce per la vendita online di

[27] Cfr. art. 25 del decreto legislativo 27 maggio 2022, n. 82.
[28] Infatti, l'art. 2 del decreto legislativo 27 maggio 2022, n. 82 definisce "fornitore di servizi" una «persona fisica o giuridica che fornisce un servizio sul mercato dell'Unione o si offre di fornire tale servizio ai consumatori nell'Unione» e "fabbricante" «una persona fisica o giuridica che fabbrica un prodotto oppure lo fa progettare o fabbricare e lo commercializza apponendovi il proprio nome o marchio d'impresa».

biglietti, pubblicazioni, merchandising; servizi di accesso a contenuti audiovisivi come piattaforme streaming di mostre o concerti; e-book reader e software per la lettura digitale; servizi bancari e di pagamento online legati all'offerta culturale ecc.). In questi casi, il sito web o l'app che veicola tali servizi digitali culturali dovevano essere resi conformi ai requisiti di accessibilità entro il 28 giugno 2025.

L'unica deroga totale nei confronti degli obblighi previsti dall'European Accessibility Act è riservata alle microimprese[29] mentre le Piccole e Medie Imprese (PMI)[30], possono derogare solo a determinate condizioni, cioè invocando "l'onere sproporzionato". Si tratta di una sorta di clausola di salvaguardia[31] che indica i casi in cui i costi per rendere accessibile un prodotto o un servizio siano tali da renderne antieconomica la produzione, distribuzione o vendita[32]. La deroga è valida anche quando l'applicazione di tutte le norme di accessibilità comportino uno stravolgimento sostanziale della natura di un prodotto o di un servizio per renderlo accessibile. Si noti che, mentre le microimprese sono completamente esentate dall'applicazione dell'European Accessibility Act – dal momento che la dimostrazione dell'onere sproporzionato sarebbe di per sé già un onere sproporzionato – le PMI devono giustificare con dati e motivazioni l'eventuale onere "sproporzionato" di messa a norma dei prodotti e servizi. È bene tenere presente che in caso di mancato rispetto del decreto sono previste delle sanzioni amministrative pecuniarie che tengono conto di vari fattori, come l'entità della non conformità, il numero delle unità di prodotti o di servizi non conformi nonché il numero degli utenti coinvolti, e che possono arrivare fino a 40.000 euro[33].

6. L'impatto dell'European Accessibility Act sul settore culturale

Come già anticipato, l'European Accessibility Act introduce obblighi e opportunità di rilievo anche per gli istituti culturali e i luoghi della cultura. Infatti, gli ambiti di applicazione della direttiva, insieme ai principi generali che essa enuncia, comportano obblighi – sia diretti che indiretti – per gli istituti culturali pubblici e, in diversi casi, anche per quelli privati, in particolare quando erogano servizi digitali accessibili al pubblico

[29] Le microimprese e le PMI sono definite nell'art. 3 dell'European Accessibility Act e nell'art. 2 del decreto legislativo 27 maggio 2022, n. 82: per "microimpresa" si intende «un'impresa che occupa meno di 10 persone e realizza un fatturato annuo oppure un totale di bilancio annuo non superiore a 2 milioni di euro».

[30] *Ivi*. Per "piccole e medie imprese" o "PMI" si intende «la categoria di imprese che occupano meno di 250 persone, il cui fatturato annuo non supera i 50 milioni di euro o il cui totale di bilancio annuo non supera i 43 milioni di euro, ma che non comprende le microimprese».

[31] Specificata all'articolo 14 del dell'European Accessibility Act e nell'art. 13 del decreto legislativo 27 maggio 2022, n. 82.

[32] Si presti attenzione al fatto che l'onere sproporzionato va documentato e la relativa documentazione conservata per un periodo non inferiore a 5 anni, inoltre la valutazione per onere sproporzionato deve essere rinnovata quando il servizio è modificato, quando è richiesto dall'autorità di controllo, e in ogni caso ogni 5 anni. Cfr. art. 13, co. 5 del decreto legislativo 27 maggio 2022, n. 82.

[33] Cfr. art. 24 del decreto legislativo 27 maggio 2022, n. 82: «salvo che il fatto costituisca reato e fuori dai casi di esenzione di cui agli articoli 3, comma 3, e 13, comma 1, l'operatore economico che contravviene alle disposizioni di cui agli articoli 3, 6, commi da 1 a 8, 8 commi da 1 a 7, 9, commi da 1 a 4, 10 e 12, commi da 1 a 4, è punito con la sanzione amministrativa pecuniaria da 5.000 euro a 40.000 euro, tenendo conto dell'entità della non conformità, del numero delle unità di prodotti o di servizi non conformi nonché del numero degli utenti coinvolti».

MODUS

che rientrano nelle categorie previste. Inoltre, l'applicazione congiunta dell'European Accessibility Act e della direttiva UE 2016/2102 delinea un quadro normativo ormai pressoché completo e comporta che gli istituti culturali, salvo poche eccezioni legate al contesto privato, debbano conformarsi agli obblighi previsti in materia di accessibilità[34]. Per comprenderne meglio la portata, si considerino i seguenti ambiti applicativi che fanno riferimento al contesto delle istituzioni culturali pubbliche (ma che in molti casi sono mutuabili anche al contesto privato).

Accessibilità dei servizi istituzionali digitali. Uno dei principali ambiti di applicazione della normativa riguarda l'accessibilità dei servizi digitali istituzionali offerti all'utenza da archivi, biblioteche e musei, – come quelli disponibili attraverso i siti web e le applicazioni per dispositivi mobili – che devono poter essere pienamente fruibili dalle persone con disabilità sensoriali, motorie o cognitive, nel rispetto delle linee guida WCAG 2.1, almeno al livello AA[35]. Questo significa, ad esempio, che i siti devono avere una struttura semantica corretta, presentare testi alternativi per le immagini, essere compatibili con gli screen reader e navigabili anche da tastiera. Analogamente, le applicazioni mobili – come quelle per la prenotazione di libri, visite o sale studio – devono offrire interfacce accessibili anche a chi ha disabilità visive, uditive, motorie o cognitive. Ad esempio, un archivio che mette a disposizione dell'utenza servizi digitali come portali web per la consultazione da remoto dei fondi, strumenti di ricerca e banche dati archivistiche online ecc. è tenuto ad assicurarne l'accessibilità da parte delle persone con disabilità prevedendo la possibilità di ingrandire i caratteri, aumentare il contrasto visivo, utilizzare la modalità di lettura facilitata o l'audio-lettura. Una biblioteca che rende disponibile una applicazione mobile per la prenotazione di libri o per l'accesso in sala studio deve accertarsi della conformità alle linee guida WCAG 2.1, almeno al livello AA. Anche nella progettazione dei sistemi di accesso online ai cataloghi (come gli Opac) deve essere tenuto in considerazione l'aspetto dell'accessibilità: essi devono essere usabili da tutti, con funzionalità compatibili con le tecnologie assistive. Una digital library deve garantire che il sito web e la piattaforma di accesso siano pienamente accessibili in maniera tale che gli utenti con disabilità possano navigare, ricercare, accedere ai contenuti e utilizzare tutte le funzioni (consultazione, download, prestito, prenotazioni ecc.) utilizzando tecnologie assistive (screen reader, comandi vocali, utilizzo della tastiera ecc.). Inoltre, deve garantire la conformità dei propri e-book e dei lettori digitali ai requisiti di accessibilità: essi devono poter essere fruibili anche da persone con disabilità, ad esempio grazie alla presenza di testi strutturati semanticamente; de-

[34] Permangono, tuttavia, alcune ambiguità interpretative, dovute in parte a traduzioni dall'inglese all'italiano non sempre aderenti al senso tecnico-giuridico originario, le quali si riflettono anche nel decreto legislativo di recepimento. Questo quadro sarà ulteriormente definito con l'adozione del decreto relativo ai prodotti da parte del Ministro delle imprese e del made in Italy, e delle *Linee guida in materia di servizi da parte dell'Agenzia per l'Italia Digitale*, come previsto, rispettivamente, dagli art. 18 e 21 del decreto legislativo 27 maggio 2022, n. 82.

[35] Come già previsto dalla direttiva UE 2016/2102 nel caso dei soggetti pubblici.

scrizioni alternative per le immagini; possibilità di personalizzazioni dell'aspetto del testo (dimensione dei caratteri, colore, contrasto); presenza di metadati per l'accessibilità[36].

Accessibilità dei servizi informativi digitali. Anche i servizi informativi digitali, come avvisi, comunicazioni, informazioni sugli orari di apertura, moduli di contatto, inviti digitali, bandi per offerte di collaborazioni, sistemi per le richieste di documenti e comunicazioni digitali in generale devono essere resi accessibili a tutti, comprese le persone con disabilità visive o uditive, e rispettare criteri di fruibilità e compatibilità con le tecnologie assistive. Ad esempio, una biblioteca che ha attivato un servizio di newsletter deve inviare le e-mail in un formato elettronico accessibile, evitando layout complessi o testo codificato come immagini. Un museo che intende pubblicare online moduli per la ricerca di volontari e la partecipazione a bandi dovrebbe utilizzare formati accessibili quali il PDF/UA[37] o l'HTML accessibile[38]. Un archivio che pubblica online il regolamento per l'accesso

[36] Il formato di elezione per la produzione di e-book accessibili è l'ePub3 (Electronic Publication 3). Sviluppato dal World Wide Web Consortium (W3C) attraverso il gruppo di lavoro del Publishing@W3C, si basa su tecnologie web come HTML5, CSS e JavaScript. Uno dei suoi punti di forza è il robusto supporto all'accessibilità, che lo rende lo strumento più adatto a soddisfare i requisiti imposti dall'European Accessibility Act.

[37] Sul versante dei formati elettronici in grado di assicurare i requisiti di accessibilità, una delle proposte più interessanti è il PDF/UA (PDF/Universal Accessibility), un profilo del formato PDF che risponde alle raccomandazioni di accessibilità ai contenuti da parte delle persone con disabilità, in conformità con le normative attuali. Affinché un documento in formato PDF possa dirsi conforme alle specifiche del formato PDF/UA deve possedere tutta una serie di caratteristiche; senza addentrarci in dettagli tecnici eccessivi, è possibile citare, a titolo di esempio, il fatto che i *tag* devono rappresentare correttamente le strutture semantiche del documento (titoli, elenchi, tabelle ecc.); sono vietati i contenuti problematici, compresi i titoli illogici, l'uso di colori/contrasti per trasmettere informazioni, codice JavaScript inaccessibile e altro ancora; le immagini devono includere descrizioni di testo alternative; le impostazioni di sicurezza devono consentire alle tecnologie assistive di accedere ai contenuti; i font devono essere incorporati e il testo deve essere mappato in Unicode ecc. Tale formato è stato definito per la prima volta nel 2012, con la pubblicazione della norma ISO 14289-1:2012 (ora ritirata) che ha introdotto il PDF/UA-1; successivamente nel 2014 è stata pubblicata la norma ISO 14289-1:2014, che conteneva una serie di correzioni rispetto alla prima versione. L'interesse verso il formato è cresciuto notevolmente solo negli ultimi anni, da una parte per il fatto che recentemente è stata pubblicata la norma ISO 14289-2:2024 che ha definito la seconda versione del formato (PDF/UA-2); dall'altra per il fatto che la scadenza del 28 giugno 2025 ha spinto le organizzazioni alla ricerca di un formato elettronico in grado di soddisfare pienamente i requisiti previsti dall'European Accessibility Act. Cfr. International Standardization Organization (ISO), *ISO 14289-1:2014 — Document management applications — Electronic document file format enhancement for accessibility — Part 1: Use of ISO 32000-1 (PDF/UA-1)*, <https://www.iso.org/standard/64599.html>; International Standardization Organization (ISO) (2024). *ISO 14289-2:2024 — Document management applications — Electronic document file format enhancement for accessibility — Part 2: Use of ISO 32000-2 (PDF/UA-2)*, <https://www.iso.org/standard/82278.html>.

[38] Per "HTML accessibile" si intende una modalità di scrittura del codice HTML che consente a tutte le persone, comprese quelle con disabilità (visive, uditive, motorie, cognitive ecc.), di accedere, comprendere e interagire efficacemente con i contenuti di una pagina web. Ad esempio, gli elementi HTML (come <header>, <nav>, <main>, <article>, <section>, <footer> per strutturare il contenuto; <h1>…<h6> per titoli e sottotitoli in modo gerarchico; , <ol>, <li> per le liste; <button>, <form>, <label>, <input> per i moduli) devono essere utilizzati secondo la loro funzione per aiutare gli screen reader e gli altri ausili tecnologici a interpretare correttamente la pagina; tutti gli elementi interattivi (link, bottoni, moduli) siano raggiungibili con il tasto TAB; i contenuti non testuali devono essere accompagnati da alternative testuali utilizzando l'attributo "alt" per descrivere il contenuto delle immagini (<img alt="Descrizione">) e fornendo trascrizioni per audio e video; ecc.

MODUS

e la consultazione in formato PDF/UA o HTML accessibile, risponde correttamente ai requisiti di accessibilità[39].

Accessibilità dei contenuti culturali digitalizzati. Sempre più spesso le istituzioni culturali digitalizzano e rendono disponibile online il proprio patrimonio: fondi archivistici, periodici, manoscritti, collezioni d'arte ecc. Tuttavia, la mera acquisizione dei documenti analogici produce tipicamente documenti digitali in formato immagine (ad esempio in formato JPEG o PDF bitmap) non leggibili mediante tecnologie assistive, come gli screen reader. Per rispettare i criteri di accessibilità, nel caso di immagini digitali ottenute dall'acquisizione di documenti cartacei, esse devono essere sottoposte a processi di riconoscimento ottico della scrittura, sia a stampa (OCR) che manoscritta (HCR)[40], e accompagnate dalla produzione dei relativi metadati, in maniera da produrre testo che sia accessibile anche da parte di utenti con disabilità visive o cognitive[41]; le immagini che non derivano da documenti testuali, i grafici e le mappe devono disporre di descrizioni alternative. Le interfacce dei portali devono consentire la consultazione anche a chi ha disabilità sensoriali.

Accessibilità dei contenuti culturali multimediali. I contenuti multimediali – come audioguide, videoguide, podcast, mostre virtuali e visite immersive – costituiscono una componente sempre più rilevante dell'offerta culturale. In base alla normativa sull'accessibilità, tali contenuti devono essere fruibili in tutte le loro parti, anche attraverso modalità alternative. Ciò implica, ad esempio, l'obbligo di corredare i contenuti video con sottotitoli sincronizzati, traduzioni in Lingua dei segni italiana (LIS), trascrizioni testuali per le persone sorde, oppure audio descrizioni per quelle cieche o ipovedenti. Le interfacce che veicolano tali contenuti devono essere semplici, intuitive e, se necessario, dotate di linguaggio semplificato per utenti con disabilità cognitive. Una mostra che utilizza QR code per l'accesso a contenuti inclusivi, oppure un museo che offre visite guidate virtuali con sottotitoli, traduzione nella LIS e audio descrizioni, rappresentano esempi concreti

[39] Ad esempio, un museo che offre un tour virtuale su una piattaforma web dovrà fornire descrizioni audio per le immagini, sottotitoli per il video e garantire che il sito sia navigabile mediante uno screen reader. Una app per dispositivi mobili per prenotare visite guidate dovrà essere utilizzabile anche da persone con disabilità visive o motorie. Un chiosco digitale in loco (per informazioni o biglietteria) dovrà includere comandi vocali o supporti tattili.

[40] I processi di riconoscimento ottico della scrittura permettono di convertire immagini contenenti testo in contenuti testuali digitali ricercabili ed elaborabili. Il riconoscimento ottico dei caratteri a stampa (OCR, Optical Character Recognition) viene utilizzato per trasformare documenti stampati o dattiloscritti (come libri, articoli, documenti amministrativi) in testi digitali modificabili. Il riconoscimento della scrittura manoscritta (HCR, Handwritten Character Recognition) si basa su una tecnologia più complessa, poiché deve interpretare le grafie individuali e irregolari presenti nei documenti manoscritti.

[41] Sulle modalità di creazione di documenti digitali accessibili si rimanda alle numerose guide che sono state pubblicate negli ultimi anni, compresa quella resa disponibile dall'Agenzia per l'Italia Digitale (AgiD) già nel 2017. Cfr. Agenzia per l'Italia Digitale, *Guida pratica per la creazione di un documento accessibile. Versione 27 marzo 2017 - Aggiornamento del documento del 18 luglio 2016*, attualmente disponibile sul sito dell'Università di Macerata, <https://www.unimc.it/it/amministrazionedigitale/accessibilita/guida-pratica-per-la-creazione-di-un-documento-accessibile/guida-pratica-per-la-creazione-di-un-documento-accessibile-1.pdf>.

MODUS

di applicazione dei principi di accessibilità. Anche le esperienze immersive e in realtà aumentata, se rivolte al pubblico, devono rispettare i requisiti di accessibilità. Ad esempio, un sito archeologico può offrire visite in realtà virtuale (VR) dotate di navigazione assistita, comandi vocali e versioni alternative in formato video o testuale. Allo stesso modo, una biblioteca universitaria che pubblica video tutorial per gli studenti deve garantire la piena fruibilità dei contenuti da parte di tutti, attraverso interfacce accessibili, audio con trascrizioni sincronizzate e compatibilità con le tecnologie assistive.

Accessibilità delle interfacce fisiche e delle postazioni self-service. L'accessibilità digitale non riguarda solo i contenuti online, ma si estende anche ai dispositivi interattivi fisici presenti nei luoghi della cultura, come totem informativi, chioschi per la biglietteria automatica, postazioni di consultazione, casse self-service e terminali per audioguide. Tali apparecchiature devono essere progettate per garantire l'utilizzo anche da parte di persone con disabilità motorie, visive o uditive, senza ostacoli né barriere operative e nel rispetto di specifici criteri di accessibilità. Le interfacce devono prevedere comandi alternativi (vocali o tattili), testi ingrandibili, feedback sia acustici che visivi, e un posizionamento conforme agli standard ergonomici. Un museo, ad esempio, che utilizza totem informativi deve dotarli di interazioni tattili, audiodescrizioni e regolazione del contrasto visivo. Un archivio che mette a disposizione postazioni per la consultazione autonoma di documenti digitalizzati deve prevedere comandi vocali e interfacce inclusive. Una biblioteca universitaria che mette a disposizione una mappa interattiva per l'orientamento deve assicurarne la fruizione anche da parte di utenti ciechi, ad esempio attraverso riproduzioni sonore o l'uso di una barra braille. Analogamente, un museo che offre un'app per guidare il visitatore nel percorso espositivo è tenuto a garantirne l'accessibilità tramite supporto vocale, comandi tattili e interfacce ad alto contrasto cromatico.

Accessibilità dei servizi commerciali e di biglietteria. L'obbligo di accessibilità si estende anche ai servizi commerciali online collegati al settore culturale, quali la vendita di riproduzioni digitali, il merchandising museale, le donazioni online, l'acquisto di biglietti d'ingresso e altri servizi similari. Quando tali attività sono gestite attraverso piattaforme digitali, esse rientrano nell'ambito di applicazione della normativa in materia di accessibilità. Pertanto, tali servizi devono essere pienamente accessibili a tutte le categorie di utenti, comprese le persone con disabilità. Ad esempio, un museo è tenuto a garantire che la propria piattaforma per l'acquisto dei biglietti sia accessibile anche agli utenti che utilizzano tecnologie assistive, come screen reader o dispositivi di navigazione da tastiera. Se offre visite virtuali a pagamento, deve assicurarsi che tali contenuti siano fruibili anche da persone con disabilità. Analogamente, un archivio che vende riproduzioni digitali dei propri documenti deve rendere accessibile sia il portale sia il processo di acquisto. Un sito di e-commerce collegato a un istituto culturale che offra gadget o pubblicazioni deve garantire un'esperienza d'acquisto priva di barriere, compatibile con

gli standard di accessibilità digitale. Anche le piattaforme per le donazioni online gestite da fondazioni culturali devono essere progettate per essere fruibili da utenti con disabilità visive, uditive, motorie o cognitive, assicurando un'interazione semplice, chiara e accessibile.

Accessibilità degli eventi culturali online. Le disposizioni in materia di accessibilità si estendono anche ai contenuti culturali audiovisivi trasmessi in streaming, sia in diretta che on demand. Rientrano in questa categoria presentazioni, conferenze, corsi, laboratori ed eventi online di vario genere. Tali iniziative devono essere accompagnate da adeguate misure di accessibilità, quali sottotitoli sincroni, trascrizioni testuali o la presenza di interpreti LIS integrati nello streaming. Ad esempio, una biblioteca che trasmette in diretta su YouTube la presentazione di un libro è tenuta a garantire l'accessibilità dell'evento, ad esempio mediante sottotitoli in tempo reale e/o trascrizioni accessibili.

Obblighi indiretti. L'European Accessibility Act comporta anche obblighi indiretti, in particolare nell'ambito dei contratti e delle forniture. Sebbene gli obblighi di accessibilità si applichino direttamente agli operatori economici che immettono sul mercato i prodotti e i servizi elencati dalla normativa europea (quali software, e-reader, terminali ecc.), essi hanno ricadute anche sugli istituti culturali che acquistano o utilizzano tali prodotti e servizi. Di conseguenza, ogni volta che un istituto culturale commissiona la realizzazione di un sito web, di un'applicazione mobile, di una mostra virtuale, di una banca dati online o di un intervento di digitalizzazione, è fondamentale che nei capitolati tecnici siano incluse clausole specifiche relative all'accessibilità[42]. Ciò risulta particolarmente rilevante per gli enti pubblici o a partecipazione pubblica, i quali sono vincolati anche dalle disposizioni della direttiva UE 2016/2102 sull'accessibilità dei siti web e delle applicazioni mobili degli enti pubblici.

7. Conclusioni

L'European Accessibility Act rappresenta una svolta epocale nel panorama normativo europeo, introducendo una serie di requisiti e standard volti a garantire l'accessibilità universale dei servizi e dei prodotti, con un impatto rilevante in molti settori, incluso quello dei beni culturali. La scadenza del 28 giugno 2025 costituisce un passaggio cruciale, che impone a istituzioni pubbliche e operatori economici non solo un adeguamento tecnico, ma anche un cambiamento culturale profondo. L'accessibilità cessa di essere un adempimento formale per divenire principio strutturale delle politiche di innovazione, inclusione e partecipazione. Per gli istituti culturali, l'entrata in vigore dell'European Accessibility Act rappresenta al contempo un obbligo e un'opportunità: da un lato, l'ob-

[42] Come previsto anche dal decreto legislativo 31 marzo 2023, n. 36 *Codice dei contratti pubblici in attuazione dell'articolo 1 della legge 21 giugno 2022, n. 78, recante delega al Governo in materia di contratti pubblici*. Cfr. l'art. 128 "Servizi alla persona": «L'affidamento deve garantire la qualità, la continuità, l'accessibilità, la disponibilità e la completezza dei servizi, tenendo conto delle esigenze specifiche delle diverse categorie di utenti, compresi i gruppi svantaggiati e promuovendo il coinvolgimento e la responsabilizzazione degli utenti».

bligo di progettare servizi digitali culturali conformi ai requisiti normativi; dall'altro, l'opportunità di rafforzare la propria missione educativa e democratica, rendendo la cultura realmente fruibile da tutti. La piena attuazione del principio di accessibilità si configura, così, non come un mero adempimento tecnico ma come un elemento costitutivo di una visione più ampia, orientata alla giustizia sociale e alla partecipazione.
In definitiva, l'European Accessibility Act interpella il settore culturale a scegliere tra una conformità formale e un impegno autentico verso un futuro in cui ogni barriera – fisica o digitale – diventa occasione per ripensare servizi, strumenti e priorità. In questo contesto, l'accessibilità non può più essere considerata un semplice obbligo normativo, ma assume il ruolo di paradigma fondativo per la costruzione di una società più equa, partecipativa e realmente inclusiva.

Starting from June 28, 2025, the European Accessibility Act (EU Directive 2019/882) introduced significant digital accessibility requirements for cultural institutions, both public and private. After providing a regulatory and conceptual framework on the topic of digital accessibility, the paper traces the evolution of European and national legislation, focusing on the interaction between the new directive and the previous directive 2016/2102. It then analyzes the direct and indirect implications for cultural institutions, distinguishing between certain areas of application. It highlights how accessibility should not be considered a mere technical or regulatory constraint, but rather an opportunity to strengthen the democratic and inclusive role of cultural institutions. In a context that is increasingly attentive to disability and social inclusion, the directive, despite presenting some interpretative challenges, offers a unique opportunity to rethink accessibility as a design paradigm and democratic mission. The essay concludes by emphasizing the urgency of cultural adaptation, aimed at fully integrating accessibility into the digital and inclusive strategies of cultural sites.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

Il museo nell'era digitale: tecnologie sensibili per una umanità aumentata

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00141

Claudia Pecoraro
Museologa e curatrice

Il museo contemporaneo è chiamato a confrontarsi con un pubblico sempre più digitalizzato, ma anche sempre più eterogeneo per età, competenze e aspettative. Questo articolo esplora il ruolo delle tecnologie digitali nei processi di narrazione museale, evidenziando l'importanza di un approccio critico e consapevole, mai invasivo e centrato sull'esperienza del visitatore. Attraverso alcuni casi esemplari – M9 di Mestre, il Museo Laboratorio della Mente, il Piccolo museo del diario, il progetto in realtà aumentata di NuvolaProject – si mostra come il digitale possa rendere visibile l'invisibile, generando esperienze più o meno empatiche e accessibili, senza sostituire ma integrando l'apparato museologico tradizionale. Il museo riafferma il suo ruolo quale luogo di relazione, immaginazione e consapevolezza, in cui la presenza delle tecnologie digitali può coniugare autorevolezza scientifica e sensibilità narrativa. Si propone infine un uso attento e sostenibile delle tecnologie, orientato a un ecosistema museale più aperto, accessibile e profondamente umano.

Premessa

Così come il digitale ha modificato in modo irreversibile le nostre vite, creando una ibridazione nella quale l'analogico e l'online coesistono in maniera fluida e si integrano senza soluzione di continuità nella nostra esistenza quotidiana, anche il mondo dei musei, che ha sempre assorbito le istanze del presente, è stato inevitabilmente investito da questa rivoluzione "*on life*".

Anche se l'accoglimento delle tecnologie nei musei non è sempre stato al passo della loro velocissima evoluzione – spesso a causa di difficoltà oggettive legate alle contingenze ma anche per una naturale resistenza timorosa di snaturarne l'essenza originaria – il digitale ha espanso il concetto di museo ben oltre i suoi perimetri fisici. Il digitale ha contribuito a superare i confini dell'esperienza culturale museale, che non si limita più al tempo della visita ma coinvolge la persona prima, durante e dopo di essa ad un grado molto più alto rispetto al passato.

Come ben racconta Nicolette Mandarano nel suo recente e brillante volume *Il digitale per i musei*[1], dalla prenotazione online del biglietto d'ingresso all'uso dei social media, dai contenuti di approfondimento sul sito web fino alle app e ai podcast, le tecnologie digitali permeano ogni aspetto della vita museale, investendo la conservazione, la

[1] Nicolette Mandarano, *Il digitale per i musei. Comunicazione, fruizione, valorizzazione*, Roma: Carocci editore, 2024.

MODUS

catalogazione e l'archiviazione, la valorizzazione, la comunicazione e la fruizione del suo patrimonio, sia esso tangibile o intangibile. Questo articolo si concentrerà, tuttavia, esclusivamente sulla presenza delle tecnologie digitali nello spazio fisico del museo, analizzando come esse possano facilitare e arricchire l'esperienza della visita a condizione di un loro uso non incontrollato ma piuttosto accorto. Non saranno presi in considerazione, in questa sede, i siti e i parchi archeologici, luoghi in cui l'applicazione delle tecnologie digitali può risultare particolarmente ampia e significativa, in particolar modo per la ricostruzione dei contesti originari, ma che richiedono riflessioni specifiche e differenti. L'intento di questo articolo è semplicemente quello di offrire un contributo alle riflessioni già in corso e formalizzate da altri studiosi ed esperti del settore[2]. Non si propone, naturalmente, come trattazione esaustiva – tanto per i limiti di spazio quanto per la vastità e la continua evoluzione della materia – ma come spunto ulteriore per approfondire un tema in rapida e costante trasformazione.

Oltre le parole d'ordine: tra abusi semantici e derive operative

Un museo contemporaneo *deve* essere multimediale, interattivo e innovativo. *Deve*, insomma, usare le tecnologie digitali. Un vero e proprio imperativo categorico, pare, dilagante da almeno un decennio, a partire da bandi e avvisi pubblici concepiti per destinare finanziamenti a grandi e piccoli musei purché le loro progettualità prevedano le tre caratteristiche sopracitate: multimedialità, interattività, innovazione.
Prima di addentrarci nello specifico di come queste istanze si siano tradotte concretamente nelle politiche museali e nei progetti realizzati, se è vero che – come ci insegnano i sociolinguisti – le parole formano il pensiero, vale la pena soffermarci proprio su queste tre, senz'altro abusate, che sono state spesso causa di fraintendimenti, di equivoche relazioni semantiche, nonché di conseguenze talvolta infauste nelle pratiche museali.

Cominciamo dalla parola "multimediale". Si tende a dimenticare che anche il più tradizionale dei musei lo è: difficilmente viene alla mente un museo monomediale. Quando a un quadro appeso a una parete viene affiancato il classico cartellino con la didascalia pertinente, siamo già dinnanzi a due tipi di medium. Già in tempi antecedenti all'avvento del digitale, nella museografia ci si poneva la questione del decadimento dell'attenzione e della memoria in seguito a interferenze di stimoli da medium di varia natura. Fin dalle origini della storia del museo pubblico, infatti, il caso più comune è un disturbo di natura verbale e visivo: nelle sale espositive sono spesso compresenti messaggi verbali scritti (i testi), che vengono elaborati dai visitatori in modalità analitico-ricostruttiva, e messaggi visivi (le opere) elaborati anche in modalità senso-motoria. Impegnando così

[2] Solo per citare le ultime pubblicazioni italiane più significative, oltre a N. Mandarano, *Il digitale per i musei*, (cit.), vale la pena di menzionare: Giuliano Gaia, *Il museo immediato. Digitale per la cultura: da Arpanet all'intelligenza artificiale*, Milano: Editrice Bibliografica, 2024; Maria Elena Colombo, *Musei e cultura digitale. Fra narrativa, pratiche e testimonianze*, Milano: Editrice Bibliografica, 2020. Si segnalano anche le ricerche tematiche ancora in corso di D\Tank, il think tank del Dipartimento di Design del Politecnico di Milano: <https://www.dtank-design.polimi.it/home-think-tank-del-dipartimento-di-design/digital-museums/>.

MODUS

diversamente il cervello, può presentarsi il rischio che i due messaggi si escludano l'un con l'altro: si crea un'interferenza, per l'appunto, un corto circuito[3]. A volte il rischio è persino più alto, quando cioè ci si trova dinnanzi a quelle «compresenze invadenti» di cui parlava già Ragghianti, ovvero nel caso in cui persino le opere esposte "si disturbano" l'una con l'altra[4]. Quando a tale stratificazione di stimoli si aggiungono quelli digitali, è essenziale considerare questi presupposti per evitare che l'arricchimento si trasformi in sovraccarico percettivo o disorientamento esperienziale.

Cosa intendiamo, invece, esattamente con il termine "interattività"? Persone che interagiscono con le cose, o persone stimolate a interagire con altre persone attraverso le cose?
Quando guardo un dipinto, sono già immersa in un processo interattivo: l'opera agisce su di me, plasma il mio sguardo, orienta il mio pensiero. Allo stesso tempo, io attribuisco senso, attivo un percorso interpretativo, porto la mia esperienza dentro l'immagine. Si attua insomma una forma di reciprocità silenziosa ma intensa, che modifica entrambi i poli dell'incontro – se non nel piano fisico, almeno su quello percettivo ed energetico. Non è forse questa, già, una delle più alte forme di interazione? Eppure, nel contesto delle nuove tecnologie applicate ai musei, questa idea sembra non essere sufficiente. L'interazione "invisibile" o "intangibile" viene spesso trascurata a favore di un modello più immediato, calcolabile e – soprattutto – gratificante nel breve termine (nel lungo termine, chissà): schermi da toccare, sensori da attivare, meccanismi che producono un effetto visibile, misurabile, documentabile.
A dispetto della preziosa, romantica premessa di Krzysztof Pomian, secondo cui *gli oggetti nei musei ci legano all'invisibile*[5], l'intangibile fatica a trovare spazio nell'orizzonte digitale dominante, che sembra preferire segnali chiari e feedback concreti, piuttosto che evocazione, riflessione, nutrimento dello spirito.

Passiamo all'ultima parola tra quelle da prendere in esame. Inseguire l'"innovazione" a tutti i costi rischia, talvolta, di appannare la possibilità di usare soluzioni comunicative, magari più semplici o già consolidate, che sarebbero assai più efficaci per gli scopi specifici di un museo. Quel che percepiamo come innovativo, del resto, non è mai una ideazione creata dal nulla: è sempre frutto di una intuizione che affonda le radici in pratiche, idee e conoscenze precedenti. L'innovazione non esiste, ha detto provocatoriamente qualcuno.

---

[3] Francesco Antinucci, *Comunicare nel museo*, Roma-Bari: Laterza, 2004, p. 141.

[4] Carlo L. Ragghianti. *Arte, fare e vedere*. Firenze: Vallecchi, 1974.

[5] Krzysztof Pomian, *Collezione e museo. Nascita di un'istituzione culturale*, Milano: Electa, 1987. In particolare, Pomian sostiene che gli oggetti nei musei «sono semiotici: visibili, ma collegati all'invisibile», nel senso che rimandano a un ordine simbolico, storico, culturale che non è immediatamente presente, ma che il museo rende percepibile attraverso la mediazione dell'oggetto. Il saggio esplora il concetto di collezione come fenomeno antropologico, sostenendo che gli oggetti raccolti nelle collezioni vivono sospesi tra il visibile e l'invisibile, poiché lo sguardo che li raggiunge non è solo quello del presente, ma anche quello del futuro. Esporre oggetti, in particolare nei musei, significa affidarli allo sguardo delle generazioni future.

Allo stesso modo, anche la stessa espressione "nuove tecnologie" suona ormai obsoleta. Viviamo immersi in un ecosistema tecnologico tanto diffuso da risultare invisibile, quasi naturale. Scriviamo su tastiere invece che a mano, affidiamo la nostra memoria a fotografie digitali e social network, ci orientiamo grazie a navigatori satellitari, otteniamo notizie e contenuti tramite algoritmi, interagiamo quotidianamente con intelligenze artificiali in grado di rispondere ai nostri comandi vocali, di gestire dispositivi smart, di automatizzare la nostra quotidianità. Prendo quindi a prestito le parole di Luigi Di Corato che definiva il concetto stesso di "innovazione" e "nuove tecnologie" «una sorta di relitto dell'era pre-tecnologica retaggio di un analfabeta digitale che vede ancora qualcosa di esotico nel più comune tra gli ingredienti del presente»[6].
Per questo, anziché finanziare i musei per l'acquisto *tout court* di tecnologie "innovative", dunque, sarebbe di sicuro più sano sostenerne la ricerca e la progettazione consapevole.
L'obiettivo non dovrebbe essere "avere la tecnologia", ma comprenderne il senso, integrarla con intelligenza – e lungimiranza – alle esigenze specifiche dei contenuti, dei contesti, dei pubblici. Riflettere su quali strumenti siano realmente necessari per raccontare in maniera significativa la propria identità, in modo *site-specific* e *sense-specific*.
Poi, solo *dopo* questo processo critico, magari, perché no: inventare qualcosa di nuovo.

Mostruosità museo(il)logiche: quando il mezzo diventa il fine

Come il sonno della ragione genera mostri, anche l'inseguimento cieco dei tre requisiti oggi così tanto richiesti – multimedialità, interattività, innovazione – ha prodotto, nel tempo, non poche mostruosità museo(il)logiche. Il problema non risiede tanto nella natura di queste tipologie di strumenti – che, appunto, dovrebbero essere considerati tali: *strumenti*, non *obiettivi* – quanto nella confusione concettuale che spesso li accompagna e nell'approccio acritico con cui vengono talvolta adottati.
Non sono rari i casi in cui musei, attratti dalla legittima possibilità di accedere a nuovi finanziamenti, hanno acquistato dispositivi tecnologici semplicemente perché un bando lo consentiva o richiedeva. Tuttavia, in alcuni casi per la fretta dettata da scadenze ravvicinate, in altri per la mancanza di personale adeguatamente formato o aggiornato, in altri ancora per l'assenza di una visione curatoriale, spesso tali attrezzature sono approdate al museo senza un progetto solido che assicurasse un felice matrimonio tra dispositivo e contenuto.
Il risultato? Tecnologie inutilizzate, talvolta rimaste inscatolate nei magazzini o, peggio ancora, in bell'esposizione lungo il percorso museale come installazioni silenziose perché… spente. Forse non è un caso che il primato in questa triste direzione, negli anni passati, sia spettato ai touch screen, una delle tecnologie più precocemente adottate dai musei, nonché tra le più longeve. Sarà forse proprio per la sua prima radice etimologica

[6] Luigi Di Corato, *Studio Azzurro ovvero del museo opera d'arte*, in: *Studio Azzurro, Musei di narrazione. Percorsi interattivi e affreschi multimediali*, a cura di F. Cirifino, E. Giardina Papa e P. Rosa, Milano: Silvana Editoriale, 2011, p. 164.

latina (*digitus*, dito)[7] che il "digitale" ha trovato nel tocco il suo gesto simbolico più diffuso e riconoscibile. Eppure, se il touch screen ha avuto in passato un'aura di innovazione, oggi il suo potere suggestivo si è fortemente affievolito. È stato rapidamente assorbito nell'uso quotidiano, banalizzato e massificato da smartphone, tablet e altri dispositivi. In molti casi, più che "innovare", il suo impiego rischia di replicare, senza alcun valore aggiunto, ciò che il visitatore già sperimenta quotidianamente fuori dal museo.
In definitiva, ciò che rende virtuoso l'uso della tecnologia non è la tecnologia in sé, ma l'elaborazione culturale e progettuale che la sostiene. È la visione curatoriale che conta: la capacità di pensare prima al senso, poi allo strumento.
Ricordare che la tecnologia è un mezzo, non un fine, significa rimettere al centro il contenuto, la visione e l'intenzione comunicativa, uniche vere bussole in grado di orientare l'esperienza museale verso una reale qualità e rilevanza.
Non bisogna, inoltre, avere paura di dichiarare da principio le difficoltà legate alla sostenibilità delle tecnologie utilizzate, così come riconoscere, in una data équipe museale, l'eventuale mancanza di competenze specifiche in organico per la gestione strategica del digitale, in modo da affrontare tali ostacoli già in fase di progettazione. Se si intende abbracciare l'ingresso della tecnologia digitale al museo, occorre infatti tener presente che tale processo richiede una pianificazione di medio e lungo periodo, che coinvolge risorse economiche e umane in uno spettro multidisciplinare.
Si segnala, in questa sede, l'ottimo lavoro di riflessione che sta conducendo di recente l'ICOM Learning Center, che invita periodicamente professioniste e professionisti di diversi ambiti culturali a confrontarsi sul tema, analizzando le barriere legate all'adozione delle tecnologie digitali nei musei, portando alla luce casi studio e proponendo soluzioni possibili[8].

Tecnologia dominante e sovraccarico sensoriale: il caso del museo M9 a Mestre

Un caso significativo è rappresentato da M9, il Museo del Novecento inaugurato a Mestre nel 2018 con l'intento di conferire identità culturale a una città storicamente percepita come la terraferma "di servizio" rispetto all'inarrivabile Venezia.
Dopo una storia quasi ventennale di tentativi per avere un museo *di* Mestre, nel 2010 la Fondazione di Venezia ha indetto un concorso internazionale per una progettazione in un'area del centro cittadino rimasta a lungo preclusa alla cittadinanza. Ad aggiudicarsi la

[7] Dal vocabolario Treccani online <https://www.treccani.it/vocabolario/digitale2/>: «agg. [dall'ingl. *digital*, der. di *digit* (dal lat. *digĭtus*, dito), cifra (di un sistema di numerazione)]. In elettronica e in informatica, qualifica che, in contrapp. ad *analogico*, si dà ad apparecchi e dispositivi che trattano grandezze sotto forma numerica, cioè convertendo i loro valori in numeri di un conveniente sistema di numerazione […]».

[8] Vedi, in particolare, il tavolo tecnico "Inclusione e multichannel audience engagement", tenutosi il 22 luglio 2024: <https://www.icom-italia.org/icom-learning-center-tavolo-tecnico-inclusione-e-multichannel-audience-engagement-report-disponibile/> e il webinar del 19 dicembre 2024, "Innovazione digitale e inclusione dei pubblici: pratiche e strategie di interazione per il patrimonio culturale": <https://www.icom-italia.org/icom-learning-center-webinar-innovazione-digitale-e-inclusione-dei-pubblici-pratiche-e-strategie-di-interazione-per-il-patrimonio-culturale-19-dicembre-2024/>.

commissione è stato uno studio berlinese d'eccellenza, composto da architetti, designer, ingegneri e modellisti, con un progetto di rigenerazione urbana basato sull'integrazione tra nuove architetture e edifici restaurati, in un'ottica ecosostenibile e armonicamente inserita nel contesto urbano.

Nei primi due piani dell'edificio principale è ambientata la vera e propria esposizione che racconta in otto sezioni tematiche la storia politica, economica, sociale e culturale del Novecento italiano. Il percorso si snoda attraverso circa sessanta installazioni multimediali e interattive che includono diverse migliaia di documenti[9]. Un terzo piano è dedicato invece alle mostre temporanee.

«M9 appartiene a una nuova generazione di musei. Per la prima volta un museo racconta in modo avvincente il Novecento» si legge sulla homepage del sito[10].

L'allestimento ipertecnologico, dichiaratamente immersivo e spettacolare – costato 9 milioni di euro, su un budget complessivo di 110 milioni – ha messo in campo tutte le strumentazioni più all'avanguardia allora a disposizione: dagli immancabili touch screen alle videoproiezioni, al videomapping, fino a realtà virtuale e realtà aumentata, 3D, teche olografiche, ambienti immersivi, gaming. Tuttavia, nonostante il sicuro rigore scientifico garantito dalla collaborazione con storici, sociologi, architetti e scrittori, l'impressione dominante è che, più che il contenuto, la vera protagonista dell'esperienza museale sia proprio la tecnologia: mai invisibile ma anzi assai esibita, finisce per invadere lo spazio e sovrastare la narrazione.

Basti osservare una fotografia delle sale principali, per rendersi conto che a emergere sono più i dispositivi che le storie (Fig. 1-2). Ciascun visitatore o visitatrice, entrando in questo ambiente buio, può cogliere con un unico sguardo l'elevato numero di postazioni da esplorare, venendo soverchiati da una moltitudine di sollecitazioni simultanee: un sovraccarico sensoriale (overload) che produce proprio quella competizione attentiva di stimoli sopracitata, che andrebbe rifuggita già in fase di design. Al "Wow-effect" iniziale si sostituisce presto il disorientamento da caos visivo e la sensazione di fatica nel dover affrontare il percorso nella sua interezza.

[9] Dalla pagina Wikipedia del museo <https://it.wikipedia.org/wiki/Museo_M9#cite_ref-16>: «circa 6.000 fotografie, 820 filmati montati in una videoproiezione complessiva di oltre dieci ore, 500 oggetti iconografici (manifesti, periodici, quotidiani e materiale grigio) e 400 file audio, provenienti da 150 archivi di interesse storico (tra cui Teche Rai, Istituto Luce, Fondazione Treccani, Centro Storico Fiat, Archivi Farabola, Fondazione archivio audiovisivo del movimento operaio e democratico, Archivio storico dell'ENI e Fondazione FS Italiane)».

[10] https://www.m9museum.it/.

Figura 1. *Museo M9, Mestre, Venezia: sala dedicata all'industrializzazione dell'Italia nel XX secolo (Foto Wikimedia Commons)*

Figura 2. *Museo M9, Mestre, Venezia: sala dedicata alla trasformazione del paesaggio italiano nel XX secolo (Foto Wikimedia Commons)*

Ogni postazione richiede un tempo congruo per essere fruita a pieno e, in effetti, le misurazioni della soglia di sostenibilità sono state effettuate per ciascuna di esse. Per esempio, è noto che l'uso di un casco con visore VR solitamente non è raccomandato oltre i cinque minuti. Tuttavia, in un caso come quello di M9, il risultato di sostenibilità dell'intero allestimento è naturalmente maggiore della somma di tutte le parti. E, in un

percorso distribuito su due piani, senza alcuna luce naturale né aria, ciò può comportare diversi disagi, senza considerare poi i casi particolari di persone con disabilità cognitive o bisogni specifici.

Se la città di Mestre auspicava la nascita di un centro culturale in grado di riflettere e rigenerare la propria identità, il risultato è stato invece un museo dedicato a un tema tanto vasto e generico – il Novecento italiano – da poter essere collocato ovunque, senza un legame evidente con il territorio.

Nonostante una campagna di comunicazione iniziale di grande impatto, se nel primo mese di apertura del museo si sono contati oltre 12.000 visitatori, nei mesi successivi la media è scesa a circa 7.000 biglietti venduti[11]. La pandemia del 2020 ha fatto il resto per portare quasi al collasso finanziario l'istituzione[12], spingendola ad abbassare il costo del biglietto e a renderlo valido per l'intera giornata. Ad oggi, molti tra i visitatori che lasciano le recensioni su TripAdvisor apprezzano la possibilità di uscire a metà visita per fare una pausa, tra un piano e l'altro.

Nel 2021 il neodirettore Luca Molinari dichiarava in una intervista l'esigenza di «rigenerare […] la collezione permanente, in modo da mantenere la parte multimediale ma facendo coesistere anche una parte fisica» lamentando come «la narrazione era al servizio della tecnologia e non il contrario. Era scomparsa la narrazione ed era rimasta la tecnologia, il che è un paradosso»[13]. Il museo, oggi sotto la direzione di Serena Bertolucci, sta vivendo una fase di ripresa anche grazie alle mostre temporanee (molto meno tecnologiche) che fanno da satellite alla "collezione" digitale permanente, al coinvolgimento di numerose associazioni, alle molteplici attività per scuole e per famiglie e agli eventi di approfondimento[14]. Un segnale che evidenzia come nessuna tecnologia da sola possa sostituire il valore relazionale e che, senza un legame affettivo con una "comunità di patrimonio", nessun museo può dirsi davvero radicato.

[11] Dati desunti da «Venezia Today», notizia del 3 gennaio 2019: *Dodicimila visitatori nel primo mese di apertura di M9 a Mestre*: <https://web.archive.org/web/20190701093642/http:/mestre.veneziatoday.it/m9-mese-visitatori-museo.html>; dal sito web del Comune di Venezia, 17 maggio 2019: *Le commissioni comunali VI e IX al Museo M9: "Progetto fondamentale per tutto il territorio"*, <https://live.comune.venezia.it/it/museo-m9-mestre-visita>.

[12] Massimiliano Zane, *Riflessioni sulla crisi di M9, il museo del '900 di Mestre*, «Artribune», 11 maggio 2020 <https://www.artribune.com/professioni-e-professionisti/politica-e-pubblica-amministrazione/2020/05/crisi-m9-museo-mestre-casa-dei-tre-oci-venezia/>.

[13] Giulia Ronchi, *Cambia tutto al museo M9 di Mestre. Intervista al neodirettore Luca Molinari sulla programmazione*, «Artribune», 16 dicembre 2021 <https://www.artribune.com/professioni-e-professionisti/2021/12/museo-m9-mestre-intervista-direttore-luca-molinari-programmazione/>.

[14] Vedi l'intervista *Il bilancio di Bertolucci: in un anno diventati museo partecipato,* pubblicata il 15 gennaio 2025 in «Il Gazettino Venezia Mestre», p. 27.

Tecnologie trasparenti e narrazioni partecipate: il caso del Museo Laboratorio della Mente

Il tempo del digitale è un parametro che scorre velocissimo, una nuova unità di misura inafferrabile con cui facciamo i conti nel nostro quotidiano. Alla lentezza di adattamento e di reazione tipica delle istituzioni museali si contrappone la velocità con cui le tecnologie vengono ideate, sviluppate e rilasciate. Queste si aggiornano e si rinnovano in tempi brevissimi, spesso nell'arco di pochi mesi, mentre i musei italiani, per consuetudine, modificano gli allestimenti solo dopo diversi decenni. A ciò si aggiunge una certa resistenza al cambiamento che investe persino gli apparati tradizionali a cui il museo è ancora fortemente ancorato: supporti grafici e didascalie sembrano spesso essere concepiti come immutabili e permanenti.

Laddove si investono risorse significative per dotarsi delle tecnologie più sofisticate e all'avanguardia, la parola "obsoleto" incombe come una spada di Damocle. Come dimostra il caso emblematico di M9, un allestimento fortemente tecnologico, sebbene inizialmente "innovativo", può risultare rigido e un suo rinnovamento assai costoso.

L'unico vero antidoto a questa obsolescenza, lo ribadiamo, è fare in modo che la tecnologia non si trasformi mai nel soggetto protagonista della narrazione, ma resti sempre un mezzo discreto, al servizio del contenuto. In quest'ottica, i dispositivi devono rimanere trasparenti, ovvero integrati nello spazio museale senza imporsi visivamente, pur lasciando percepire chiaramente i loro effetti esperienziali.

A tal proposito, non si può non citare quello che probabilmente si configura come il punto di riferimento storicamente più importante in Italia, Studio Azzurro, che porta avanti questa visione come un vero e proprio vessillo, sia nella sua pratica ormai quarantennale sia a livello di teorizzazione[15]. Per il collettivo milanese, mettere in evidenza le conseguenze del dispositivo piuttosto che la tecnica stessa significa «non complicare tecnologicamente il banale» , abbassare l'apparenza e il valore simbolico delle macchine a favore della temperatura sensibile dell'ambiente, così da creare una condizione di maggiore familiarità, alla portata dei più. Ecco allora l'importanza delle cosiddette interfacce naturali: dispositivi interattivi che non richiedono la mediazione di protesi tecnologiche evidenti (visori, mouse, tastiere, bottoni…), ma si attivano con gesti spontanei, voci, sguardi, soffi, movimenti. Viene così a crearsi una condizione di maggiore naturalezza che non è frustrata da modalità non a tutti familiari. Questo approccio riduce il divario tra visitatore e strumento, attivando una dimensione comportamentale che trasforma il fruitore in co-autore del percorso museale, che a sua volta si espande al di là della forma e investe il campo della relazione. La narrazione prende quindi una strada più evocativa che informativa, coinvolge la componente emotiva dei visitatori generando uno spazio aperto in cui ogni persona può costruire in modo attivo un'esperienza personale e partecipata, diversamente da una imposta dall'alto. Per altro, questi tipi di allestimenti, anziché una spiegazione lineare e organica, prediligono di solito una narrazione particellare dei linguaggi multimediali tali da attivare sensibilità diverse e da non imporre un unico punto di vista. L'accostamento

[15] Solo per citare un esempio tra la sua ricchissima bibliografia: *Studio Azzurro, Musei di narrazione*, cit.

di documenti originali, racconti orali, gesti, immagini, suoni, crea un'accumulazione di frammenti che possono essere intesi come unità mobili di una struttura aperta.
Un linguaggio, insomma, che rispecchia l'orizzontalità e la coralità della costruzione della conoscenza a cui ci ha abituati il web, regno dei contenuti generati dagli utenti, e dunque più in linea con le esigenze di un museo che voglia abbracciare le istanze della contemporaneità. Un habitat narrativo, per di più, che incoraggia la possibilità di esperire la visita insieme agli altri, in una dinamica relazionale non solo tra la persona e le cose ma anche tra le persone stesse in una forma di socialità collaborativa.
Esempio paradigmatico di questo approccio, tra i "musei di narrazione" concepiti da Studio Azzurro, è il Museo Laboratorio della Mente di Roma[16].
Situato all'interno dell'ex manicomio di Santa Maria della Pietà, il Museo nasce come progetto della ASL Roma 1 con l'obiettivo di documentare la storia dell'istituzione manicomiale e di elaborare una costante riflessione sul paradigma salute/malattia, sull'alterità, l'inclusione, sulla politica delle cure e sul coinvolgimento della comunità. Il museo propone un percorso immersivo, multisensoriale e interattivo che stimola la riflessione sul concetto di normalità, identità e memoria collettiva. L'allestimento coinvolge il visitatore in una continua oscillazione tra elementi reali – la collezione scientifica, i dispositivi medici, i manufatti storici dell'ex manicomio, le voci degli internati, la raccolta di arte irregolare – e virtuali, stimolando la partecipazione attiva del pubblico.
L'utilizzo delle tecnologie, dalle videoinstallazioni alle interfacce tattili e sonore, è perfettamente integrato con gli ambienti originari e con i contenuti, in modo da coinvolgere emotivamente visitatrici e visitatori senza mai distrarli con l'apparato tecnico.
Chi visita il museo è chiamato a "fare esperienza" piuttosto che semplicemente osservare: può interagire con racconti orali, toccare oggetti, ascoltare voci, esplorare stanze in cui la narrazione prende corpo attraverso suggestioni visive, sonore e motorie (Fig. 3). È una visita che lascia il segno, perché non si limita a informare, ma invita a riflettere, a mettersi in relazione con l'altro da sé e con sé stessi.

Figura 3. *Museo Laboratorio della Mente, Roma. Esempi di due postazioni attivate dalle posture e dai gesti dei visitatori (Foto Studio Azzurro, per gentile concessione)*

[16] *Museo Laboratorio della Mente*, a cura di UOS Centro Studi e Ricerche ASL Roma e Studio Azzurro, Milano: SilvanaEditoriale, 2010; *Museo Laboratorio della Mente. Portatori di storie. Da vicino nessuno è normale*, a cura di UOS Centro Studi e Ricerche ASL Roma e Studio Azzurro, Milano: Silvana Editoriale, 2012, secondo catalogo dedicato all'allestimento "Portatori di storie"; <https://www.museodellamente.it/>.

In definitiva, il Museo Laboratorio della Mente incarna perfettamente quella museologia esperienziale e partecipativa che, pur facendo largo e sofisticato uso della tecnologia, la impiega con sobria delicatezza e profondità concettuale. Un approccio che apre la strada ad una modalità di fruizione culturale umana, dialogica e autentica.
Non è stato un caso aver scelto di raccontare, quale buona pratica, un museo inaugurato nel 2000: nonostante il Museo Laboratorio della Mente non si possa più dire giovanissimo, i segni di un suo invecchiamento sono davvero trascurabili. Al momento della scrittura di questo articolo, il museo è temporaneamente chiuso per i lavori di ristrutturazione del padiglione 6 e per l'ampliamento del percorso espositivo. La curiosità verso la sua riapertura è grande, soprattutto per scoprire come evolverà un progetto che, fin dalla sua nascita, ha saputo coniugare tecnologia digitale e sensibilità narrativa.

Verso un'esperienza empatica e aumentata dell'invisibile

Il museo, come ricordato in precedenza, è per eccellenza il luogo dell'invisibile: uno spazio in cui l'esperienza di coloro che lo frequentano si costruisce come una somma irripetibile e non lineare tra immediatezza, frammentazione, incontro con l'alterità, scoperta, riconoscimento (o disconoscimento), conoscenze pregresse, e il sistema delle proprie sfere sociali, culturali, affettive ed emotive, intrecciati con le narrazioni offerte.
Se il museo intende produrre un'esperienza davvero originale, al di fuori e in contrasto con l'omologazione imposta dalla mediatizzazione contemporanea, alla conoscenza deve affiancare anche l'immaginazione e l'emozione, attraverso un avvicinamento empatico con il contenuto-oggetto della sua narrazione. Rapportarsi all'invisibile significa quindi che oggi non è più sufficiente mostrare i cocci di un'anfora: è necessario far immaginare i gesti che l'hanno prodotta, utilizzata, e le vicende di cui è stata testimone. È proprio in questa direzione che l'applicazione tecnologica può intervenire e, se orientata nel modo giusto, giocare un ruolo fondamentale generando «un significativo metalivello: un campo esperienziale in cui puoi provare, scoprire, modificare questi linguaggi tecnologici in una pratica non fine a sé stessa»[17].
Un esempio particolarmente virtuoso è rappresentato da uno dei più recenti lavori del collettivo NuvolaProject, fondato da Gaia Riposati e Massimo Di Leo, specializzato in performing media e nella relazione tra fisico e digitale, tra le intelligenze umane e quelle artificiali.
In occasione della mostra "Antonio Donghi. La magia del silenzio" esposta a Palazzo Merulana[18], è stato presentato in anteprima il loro progetto artistico "Imaginaria" che ha animato in realtà aumentata alcuni dei quadri più iconici di Donghi, oltre ad altre opere di artisti italiani della prima metà del Novecento facenti parte della collezione permanente del museo. L'iniziativa non è rimasta un evento temporaneo: Palazzo Merulana

[17] *Studio Azzurro, Musei di narrazione*, cit., p. 7.
[18] Mostra a cura di Fabio Benzi, Palazzo Merulana, Roma, 9 febbraio-20 ottobre 2024. <https://imaginaria.nuvolaproject.cloud/>; <https://www.art-usi.it/a-palazzo-merulana-le-opere-si-raccontano-in-realta-aumentata-e-con-la-lingua-dei-segni/?fbclid=IwAR3BPAB7HDWeFKVrBHkUef0iq5cOOfN3Vy-sZWj0yGLDk4hPRT6FtbGwipw>.

ha scelto di integrare stabilmente l'esperienza nel percorso museale, compiendo così un passo strutturale e significativo verso una fruizione aumentata e dialogica dell'arte, che pone in relazione materia e immaterialità. Attraverso la tecnologia AR (realtà aumentata), le opere inquadrate attraverso un'applicazione – che i visitatori possono scaricare dai propri dispositivi mobili – prendono vita e performano in una vera e propria drammaturgia a cui gli autori hanno lavorato con la consulenza scientifica degli storici dell'arte del museo. Il pubblico non riceve soltanto contenuti visivi e vocali aggiuntivi ma viene invitato a interagire con l'opera esplorando lo spazio e scoprendo micro-narrazioni legate ai personaggi rappresentati, all'autore, e al clima storico-culturale del tempo. Per di più, ogni esplorazione prevede una versione recitata da Gaia Riposati in italiano, interpretata da una attrice segnante in lingua italiana dei segni (LIS), e una recitata in inglese con l'interpretazione di una attrice segnante nella lingua internazionale dei segni, garantendo anche piena accessibilità alle persone sorde (Fig. 4). Ponendo attenzione anche nella direzione dell'inclusione, l'esperienza di visita viene in questo modo "aumentata", attraverso il potere suggestivo dell'evocazione poetica e della meraviglia. La tecnologia digitale non distrae quindi dall'opera d'arte, al contrario consente di instaurare un rapporto di intimità e prossimità con il contenuto, sempre rispettandone l'aura e la filologia, per altro in dialogo con altre opere vicine e dunque in armonia con lo spazio allestitivo e la visione curatoriale. Quando il dispositivo viene spento e riposto, lo sguardo del visitatore sull'opera risulta certamente rinnovato. Lo scopo è quello di offrire un'esperienza culturale non residuale, come invece spesso accade, ma di un livello più approfondito, consapevole, intenso e gratificante.

Figura 4. *Il progetto "Imaginaria" di NuvolaProject a Palazzo Merulana, Roma: Realtà aumentata su Gita in barca di Antonio Donghi, con attrice segnante che interpreta la narrazione in lingua italiana dei segni*

Un intento analogo anima anche il Piccolo museo del diario di Pieve Santo Stefano, in provincia di Arezzo[19], che rappresenta un altro esempio virtuoso di museo capace di rendere visibile l'invisibile. Qui, un intenso percorso multisensoriale e interattivo è nato per raccontare l'Archivio Diaristico Nazionale e le preziose testimonianze autobiografiche che esso conserva. Il percorso museale conduce per mano il pubblico attraverso le scritture di persone comuni che hanno raccontato l'Italia attraverso memorie private, che da storie singole e personali sono diventate storie collettive e universali. In un museo la cui collezione è composta unicamente da parole scritte, solo la tecnologia poteva farsi strumento sensibile capace di dare voce, corpo e spazio all'intangibile densità delle storie custodite. L'intimità dei racconti autobiografici viene restituita al pubblico attraverso un allestimento multimediale, progettato dallo studio dotdotdot, che coinvolge profondamente i sensi e le emozioni. Le tecnologie non invadono mai lo spazio né sono spettacolarizzanti: strumenti narrativi al servizio della parola scritta, non sovrastano mai il potere evocativo delle parole, piuttosto lo rinforzano con una regia sensibile ed empatica. Aprendo cassetti, muovendo schedari d'archivio, avvicinandosi a oggetti, vengono evocati voci, luoghi, vissuti privati… il "fruscio degli altri"[20] (Fig. 5). Anche in questo caso, l'uso accorto del digitale consente di accedere a un livello più intimo e partecipato della fruizione museale, in cui il racconto dell'altro diventa occasione per riflettere su se stessi, generando una forma di conoscenza tanto personale quanto universale. Uno dei musei più piccoli d'Italia, come recita con orgoglio programmatico il suo stesso nome, offre insomma una visione in grande: un'esperienza immersiva, profondamente umana, che lascia un segno difficile da dimenticare.

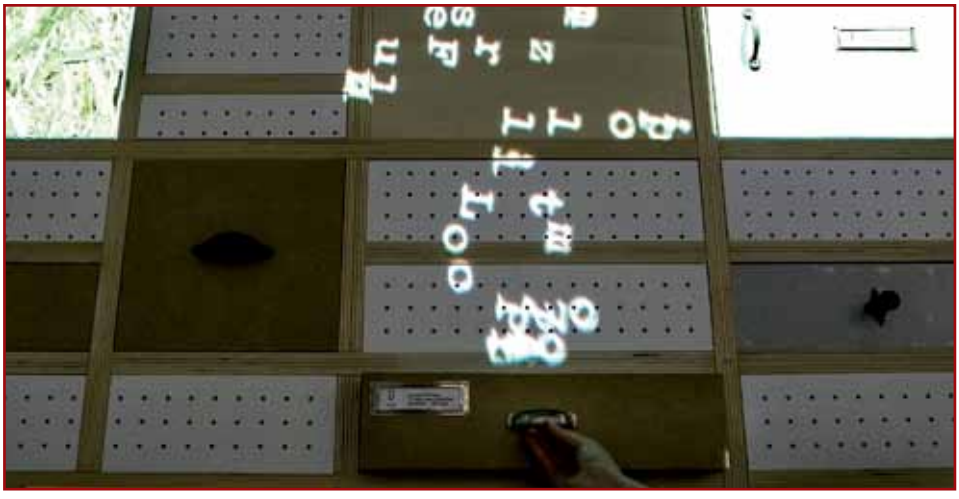

Figura 5. *Piccolo museo del diario di Pieve Santo Stefano, i cassetti della memoria (Foto Wikimedia Commons)*

[19] https://www.piccolomuseodeldiario.it/.

[20] Con questa espressione poetica, Saverio Tutino, fondatore dell'Archivio Diaristico Nazionale nel 1984, amava definire le storie narrate in diari, epistolari, memorie autobiografiche, parole private, scritti di gente comune, oggi custoditi al Piccolo museo del diario.

MODUS

Conclusioni: verso un museo digitale consapevole

Come possono i musei rispondere in modo significativo a un pubblico sempre più digitalizzato, dunque?[21] Quel pubblico che è già abituato a leggere il mondo attraverso i propri dispositivi entra al museo portando con sé tecnologia e abitudini digitali, seppure con differenti gradi di alfabetizzazione tecnologica.

Gli sforzi di introduzione del digitale nei musei si concentrano spesso nel tentativo di intercettare un pubblico genericamente "giovane", ma si trascura il fatto che la maggior parte dei visitatori attuali appartiene alla fascia over 50. Il digitale non parla ancora a tutti. Esiste ancora disparità nell'accesso e nell'utilizzo delle tecnologie digitali, un *digital divide* che richiede attenzione progettuale. Se da una parte la tecnologia apre indubbiamente molte porte sul fronte dell'accessibilità, dall'altra rischia di escludere chi non dispone degli strumenti, della formazione o della motivazione per utilizzarla. Il pubblico del museo, reale e potenziale, è numeroso e diversificato: è impossibile progettare un'esperienza sempre *tailor made*, così come lo era anche prima dell'avvento del digitale. Tuttavia, quante più possibilità si offrono, tanto più si costruisce un'esperienza culturale potenzialmente accessibile e democratica.

Il museo, in quanto luogo non solo di presentazione ma anche di interpretazione e rappresentazione, può trovare nel digitale un alleato versatile per offrire nuove chiavi di lettura, costruire significati e favorire relazioni. La rigidità delle narrazioni, il palinsesto fisso, le sale ordinatamente cronologiche sono da sempre parte delle sue ritualità. Il digitale ha il potenziale per scardinare almeno in parte tale immutabilità, aprendosi a punti di vista corali e plurali – maggiormente in linea con ciò che accade nelle nostre vite *on life*, dove la costruzione della conoscenza avviene attraverso processi sempre meno verticali e più partecipativi.

Se da una parte bisogna riconoscere che il sistema informativo tradizionale presenta dei limiti con cui far i conti[22], come l'abitudine sempre meno sviluppata e diffusa a leggere testi lunghi, il digitale rappresenta una chiave in più nel mazzo che, almeno per il momento, non sostituisce completamente gli apparati museologici classici ma li implementa. C'è chi preferisce leggere la didascalia sulla parete, chi si lascia coinvolgere da un tavolo touchscreen, chi ama lasciarsi trasportare da una voce umana (che sia in persona, in audioguida o in ologramma…) e chi desidera una fruizione libera, priva di ogni mediazione didattica o interpretativa. È compito del museo garantire comunque un'autorevolezza scientifica radicata nell'expertise del suo staff, la cui preparazione e visione deve rimanere una risorsa non negoziabile.

[21] È la domanda corretta che si pone Giuliano Gaia nel corso dell'incontro-presentazione "Musei e digitale: buone pratiche, strumenti e nuovi orizzonti" tenutosi al Salone Internazionale del Libro di Torino il 16 maggio 2025, a cui hanno partecipato Christian Greco, Nicolette Mandarano ed Emanuela Totaro: <https://www.youtube.com/watch?v=WEFhDPoOPVA>.

[22] Cfr. le considerazioni di Paola Castellucci e Elisabetta Gomelino nell'articolo *Chatbot. Un giorno, al museo*, «DigItalia. Rivista del digitale nei beni culturali», 2 (2021), p. 9-24, <https://digitalia.cultura.gov.it/article/view/2879>, che affronta opportunità, limiti e scenari futuri possibili del chatbot declinato come guida virtuale, apprendimento ludico e aumento dell'esperienza di visita.

MODUS

Il digitale può – e deve – rientrare tra gli obiettivi che si prefigge la mission del museo, come risposta ai bisogni e alle domande del presente. Va tuttavia usato in modo consapevole: il museo deve praticare un consumo critico della tecnologia, saperla selezionare, persino rifiutarla quando non necessaria, progettando allestimenti a basso impatto e prevenendo l'*e-waste*. Solo così potrà contribuire a un ecosistema digitale pulito, aperto e rigenerativo[23].

Il museo deve restare un luogo affidabile, dove, lo ribadiamo, si può incontrare l'altro da sé e ritrovare sé stessi. Questo processo richiede tempo, ascolto e profondità. È giusto, allora, che il museo continui a essere uno spazio del tempo lento, dove il sovraccarico informativo e il rumore tecnologico non trovano posto, al contrario di ciò che caratterizza la nostra vita *on life*.

Se vogliamo, e crediamo davvero, che il museo continui a essere un luogo di attraversamento consapevole, di sosta, riflessione, curiosità, approfondimento e spirito critico, dobbiamo progettare esperienze che offrano contenuti di valore, integrati con modalità di fruizione che vadano oltre lo scroll distratto dei nostri smartphone. Serve un'interazione che coinvolga il corpo, le emozioni, l'immaginazione: un incontro autentico e trasformativo con la cultura in senso ampio.

Il digitale, se impiegato con intelligenza e sensibilità, non è una minaccia, piuttosto un'opportunità: una voce in più, non l'unica. Un mezzo per rendere il museo più aperto, plurale e profondamente umano.

The contemporary museum is increasingly called upon to engage with a digitalized public, which is also more diverse in terms of age, skills, and expectations. This article explores the role of digital technologies in museum storytelling processes, emphasizing the importance of a critical and conscious approach—never invasive and always centered on the visitor's experience. Through several exemplary cases—M9 in Mestre, the Museo Laboratorio della Mente, the Piccolo Museo del Diario, and the NuvolaProject's augmented reality project—the discussion illustrates how digital tools can make the invisible visible. These technologies help generate experiences that are, to varying degrees, empathetic and accessible, not replacing but integrating traditional museological frameworks. The museum thus reasserts its role as a place of connection, imagination, and awareness, where the presence of digital technologies can successfully combine scientific authority with narrative sensitivity. In conclusion, the article advocates for a careful and sustainable use of technology—one aimed at cultivating a museum ecosystem that is more open, accessible, and profoundly human.

[23] Gerry McGovern et al. *Ecologia digitale. Per una tecnologia al servizio di persone, società e ambiente*, Milano: Altraeconomia Editore, 2022.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

Riferimenti bibliografici

Francesco Antinucci. *Comunicare nel museo*. Roma-Bari: Laterza, 2004.

Paola Castellucci — Elisabetta Gomelino. *Chatbot. Un giorno, al museo*. «DigItalia. Rivista del digitale nei beni culturali», 2 (2021), p. 9-24.

Maria Elena Colombo. *Musei e cultura digitale. Fra narrativa, pratiche e testimonianze*. Milano: Editrice Bibliografica, 2020.

Luigi Di Corato. *Studio Azzurro ovvero del museo opera d'arte*. In: *Studio Azzurro, Musei di narrazione. Percorsi interattivi e affreschi multimediali*, a cura di F. Cirifino, E. Giardina Papa e P. Rosa. Milano: Silvana Editoriale, 2011, p. 164-177.

Giuliano Gaia. *Il museo immediato. Digitale per la cultura: da Arpanet all'intelligenza artificiale*. Milano: Editrice Bibliografica, 2024.

Nicolette Mandarano. *Il digitale per i musei. Comunicazione, fruizione, valorizzazione*. Roma: Carocci Editore, 2024.

Gerry McGovern et al. *Ecologia digitale. Per una tecnologia al servizio di persone, società e ambiente*. Milano: Altraeconomia Editore, 2022.

Museo Laboratorio della Mente, a cura di UOS Centro Studi e Ricerche ASL Roma e Studio Azzurro. Milano: Silvana Editoriale, 2010.

Museo Laboratorio della Mente. Portatori di storie. Da vicino nessuno è normale, a cura di UOS Centro Studi e Ricerche ASL Roma e Studio Azzurro. Milano: Silvana Editoriale, 2012.

Krzysztof Pomian. *Collezione e museo. Nascita di un'istituzione culturale*. Milano: Electa, 1987. Edizione originale: *Collectionneurs, amateurs et curieux*. Parigi: Gallimard, 1987.

Carlo Ludovico Ragghianti. *Arte, fare e vedere*. Firenze: Vallecchi, 1974.

Studio Azzurro. *Musei di narrazione. Percorsi interattivi e affreschi multimediali*, a cura di F. Cirifino, E. Giardina Papa, e P. Rosa. Milano: Silvana Editoriale, 2011.

Vantaggi e limiti della modellazione solida applicata all'archeologia

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00142

Silvana Rapuano
Università della Campania Luigi Vanvitelli. Dipartimento di Lettere e Beni Culturali

*In questo contributo si intende riflettere sulla realizzazione e sull'uso di modelli digitali tridimensionali (3D) nella ricerca archeologica. Creare un modello 3D di un manufatto esistente (o in progetto) consente di gestire e combinare moltissime informazioni, tanto che il modello può essere considerato la controparte virtuale, il cosiddetto gemello digitale (*digital twin*), del sistema fisico rappresentato. In ambito archeologico, attraverso i gemelli digitali, è possibile l'interazione e lo studio di siti o manufatti e, ai fini della ricerca, verificare ipotesi sulle loro fasi storiche e architettoniche. Inoltre, un modello 3D può esser un valido strumento di divulgazione del patrimonio. Tali tematiche saranno sviscerate partendo da una serie di esperienze di ricostruzioni virtuali di siti di età medievale presenti nella città di Benevento, realtà urbana caratterizzata da molteplici trasformazioni e vicende insediative. Verranno illustrati il procedimento che ha portato all'attuazione di tali prodotti, i risultati ottenuti in termini di comprensione e analisi dei siti e le prospettive di valorizzazione e divulgazione dei dati scientifici, anche considerando i limiti e le criticità di tale tipo di ricerca.*

1. Introduzione

L'uso di modelli virtuali è da tempo ampiamente adottato in numerosi ambiti disciplinari, spaziando dall'architettura all'ingegneria meccanica e aeronautica, fino ad includere attività di documentazione, valorizzazione e tutela del patrimonio storico-artistico. Nell'ambito delle scienze connesse con i beni culturali, le ricostruzioni virtuali sono oggi considerate uno strumento digitale *standard*, tanto come metodo accreditato per la visualizzazione e il trasferimento di conoscenze, quanto come mezzo di supporto alla ricerca scientifica[1].

Realizzare il modello, in grafica tridimensionale (3D), di edifici o manufatti archeologici consente, infatti, oltre che una più diretta conoscenza, anche una migliore gestione degli stessi beni. Questo processo di scambio e gestione di informazioni tra l'oggetto (o il sistema) reale e la sua controparte virtuale fa sì che quest'ultima possa configurarsi come il gemello digitale (*digital twin*) del primo, per usare un termine mutuato dall'ambito ingegneristico. Tale concetto ha guadagnato un crescente interesse negli ultimi anni nell'ambito della ricerca scientifica: un gemello digitale include uno spazio reale, uno spazio virtuale e un flusso bidirezionale di dati tra di essi. In sostanza, la controparte

[1] Pfarr-Harfst — Grellert 2016, p. 39–49.

MODUS

virtuale di un oggetto o di un sistema fisico si basa sulla raccolta di dati attraverso una varietà di fonti come simulazioni, sensori e documentazione grafica[2].
I gemelli digitali, quindi, sono diventati sempre più comuni nel campo del patrimonio culturale grazie alla loro capacità intrinseca di preservare informazioni fragili o esposte al rischio di perdita. Attraverso di essi è possibile l'interazione e lo studio di siti o manufatti archeologici danneggiati, perduti, o che potrebbero essere difficili o impossibili da visitare di persona a causa di qualche tipo di restrizione (oggetti protetti da vetrate o conservati nei depositi museali, oppure inaccessibili anche temporaneamente come nel caso della chiusura senza precedenti dei musei a causa della pandemia globale di Covid-19, o anche pericolosi da raggiungere perché localizzati in aree di conflitto ecc.).
I gemelli digitali consentono anche il monitoraggio in tempo reale di grandezze fisiche (temperatura, umidità ecc.) rilevanti ai fini della tutela delle entità fisiche duplicate, se opportunamente equipaggiate di sensori[3].
Dunque, i modelli tridimensionali sono un valido strumento per la comunità scientifica, ma anche per sperimentare forme di comunicazione diretta attraverso le quali è possibile innescare un processo culturale che coinvolga il pubblico dei non addetti ai lavori.
Sono sempre più diffuse, infatti, piattaforme informatiche basate sull'integrazione tra Building Information Modeling (BIM) e Geographic Information Systems (GIS) che generano i gemelli digitali spaziali (*geospatial digital twin*) dove sono raccolti e rappresentati dati paesaggistici e archeologici. Attraverso tali database aperti e condivisi, facilmente accessibili tramite computer, smartphone o terminali video dedicati, è possibile osservare, consultare e preservare il patrimonio culturale, superando le barriere fisiche con percorsi interattivi che consentono la visita virtuale di luoghi di interesse storico e artistico[4].
In sintesi, un *digital twin* può essere visto come una raccolta di dati che evolve dinamicamente con l'aggiunta di nuove informazioni o con il cambiamento dell'oggetto di riferimento, un sistema che può essere sfruttato nei processi di apprendimento e protezione del patrimonio culturale[5].

Di particolare interesse risultano i casi di ricostruzione 3D degli assetti originari di manufatti archeologici di rilevante significato storico. È questo un settore di ricerca altamente dinamico che richiede metodologie capaci di garantire un rigoroso approccio scientifico, partendo dai rilievi delle parti sopravvissute fino alla modellazione completa. Tale rigore è fondamentale, in quanto nei reperti archeologici si riscontra frequentemente che le porzioni preservate sono di gran lunga inferiori rispetto a quelle andate perdute. Di conseguenza, le informazioni raccolte direttamente sul campo risultano spesso insufficienti per una ricostruzione accurata.

[2] Liu — Lu — Chen — Yan 2024; Niccolucci — Felicetti 2024, p. 121.
[3] Hutson — Weber — Russo 2023.
[4] Pasini 2004; Parry 2005.
[5] Babilio — Rapuano 2024.

Tale carenza di dati primari può essere compensata, integrando le informazioni con quelle derivanti dalle fonti documentarie e dalla comparazione con contesti analoghi e cronologicamente vicini. Nei casi più complessi, quando i manufatti appartengono a epoche remote con intricate vicende storiche, la ricostruzione virtuale deve necessariamente poggiare su una base dati solida e verificata.
Il modello 3D di un monumento o sito diventa così un mezzo per verificare ipotesi sulle sue fasi storiche e architettoniche. Partendo da tracce visibili, è possibile riempire virtualmente unità stratigrafiche negative, cioè superfici con evidenti segni di asportazione di materiale o di collasso, e proporre una ricostruzione che corrisponda alla lacuna. Viceversa, è anche possibile smontare le unità stratigrafiche che coprono o modificano strutture precedenti al fine di riesporre le forme originarie. Tali operazioni devono essere supportate da un'attenta analisi e da un'accurata mappatura delle stratigrafie del monumento o del sito. In tale senso il prodotto diventa una sintesi della ricerca, espressa visivamente, ed è molto più flessibile rispetto ai metodi di rappresentazione tradizionali, perché consente aggiornamenti o modifiche. Ad esempio, man mano che vengono trovate altre fonti o dati archeologici, è possibile adeguare il modello di conseguenza[6].

2. Tre esperienze di modellazione 3D

In linea con l'approccio metodologico esposto nei precedenti paragrafi, sono stati ricostruiti virtualmente i complessi archeologici dell'area dell'Arco del Sacramento, del monastero di Sant'Ilario a Port'Aurea e della chiesa di Santa Sofia, situati nell'area urbana di Benevento. Questo ha consentito di approfondire i risultati delle ricerche e di verificare le ipotesi formulate riguardo alle fasi storico-archeologiche di tali siti.
Sebbene le ricostruzioni, realizzate in anni diversi, siano state elaborate con tecnologie man mano più performanti, in tutti i tre progetti la metodologia utilizzata nell'integrazione dei dati per le ricostruzioni si è basata su:

– *interpolazione*, un processo simile al riempimento di vuoti, che consente di ricollegare due elementi distanti;
– *anastilosi*, che consiste nella ricomposizione delle parti di un monumento che devono essere ricollocati nella loro posizione originale;
– *analogia*, che consente di ricostruire un apparato coerente con l'oggetto studiato, in termini di linguaggio architettonico e stilistico, attraverso corrispondenze di similarità tra gli elementi attestati e documentati[7].

[6] *Ibidem*.
[7] Babilio — Rapuano 2023; Gabellone 2019.

MODUS

L'area dell'arco del Sacramento (Fig. 1), posta vicino all'*insula episcopalis*, è stata oggetto di scavi archeologici negli anni 2004-2007[8]. Dalle indagini sono state individuate tre macro-fasi storico-archeologiche:

– *Fase 1*. Nell'età imperiale l'area era occupata da un grande impianto termale di cui sono emersi i resti del calidarium, dei relativi impianti idraulici e di strutture pavimentali (Fig. 1a). Il complesso era posto nei pressi del foro, il cui ingresso era monumentalizzato dal cosiddetto arco del Sacramento (definito così in età moderna dal nome della via che lo attraversava).
– *Fase 2*. Durante gli scavi è stata messa in luce parte della cinta urbica del IV secolo, consistente in un tratto di muro costruito tra l'arco del Sacramento, trasformato in porta urbica e munito poi nel V secolo da una torre pentagonale, ed un secondo arco, che al momento dello scavo era inglobato in un edificio medievale (Fig. 1b)
– *Fase 3*. In età moderna l'*insula* era occupata dall'*Ospedale delle donne di San Gaetano*, che fu dismesso negli anni venti del Novecento e poi distrutto nel 1943 dai bombardamenti alleati. A lungo ricoperta dalle macerie, l'area subì notevoli sbancamenti negli anni '60 per la realizzazione di un palazzo la cui costruzione rimase interrotta (Fig. 1c).

Il lavoro di modellazione ha aiutato a fondere le molte tracce archeologiche e chiarire la complessità delle stratigrafie danneggiate anche dagli sbancamenti moderni, attraverso l'integrazione con le numerose fonti iconografiche disponibili e i rilievi di scavo eseguiti principalmente con una stazione totale, una Camera semi-metrica Rollei d 30[9] e il software PhotoMetric. La modellazione 3D è stata realizzata con Autodesk Autocad 2007[10] e la texturizzazione con Google SketchUp *6*[11], versione allora freeware.
Nei modelli elaborati al tempo si nota una certa rigidità delle forme e manca la resa fotorealistica, sebbene risultino verosimili grazie all'uso di *texture* ricavate dai rilievi fotogrammetrici o, per le superfici totalmente ricostruite, dalla biblioteca grafica del programma adattandone le dimensioni a quelle degli edifici.

[8] Le indagini nell'area dell'arco del Sacramento, anni 2004-2007, sono state condotte dall'equipe di archeologi medievisti del Dipartimento di Lettere e Beni Culturali (DiLBeC) dell'Università della Campania Vanvitelli, sotto la direzione scientifica di Marcello Rotili, in base a un accordo fra il Comune di Benevento e la Soprintendenza Archeologica di Salerno-Avellino-Benevento, nell'ambito di un progetto di riqualificazione urbana (Rotili 2006, p. 71-75; Rotili — Rapuano — Cataldo 2010, p. 309-328; Rapuano — Busino — Rotili 2016, p. 63-80).
[9] https://www.rollei.de/.
[10] https://www.autodesk.com/.
[11] Cfr. <https://archive.org/details/google-sketch-up-6_0_1099>.

MODUS

Figura 1. *Area dell'arco del Sacramento, ricostruzione 3D: (a) Fase 1, età romana; (b) Fase 2, età tardo antica-alto medievale; (c) Fase 3, età bassomedievale-moderna (S. Rapuano)*

MODUS

La chiesa di Sant'Ilario è situata in un'area cruciale della città di Benevento, in prossimità dell'arco di Traiano, riutilizzato come porta urbica dal IV secolo d.C. e definito, per questo, Port'Aurea. Fu edificata tra la fine del VII e la prima metà dell'VIII secolo e poco più tardi ad essa fu annesso un monastero benedettino. Negli anni 2000-2002 l'area venne indagata con scavi archeologici e furono riportati, sia pure non per intero, i resti del cenobio benedettino di cui si era persa memoria[12]. In base ai rilievi fotogrammetrici e ai disegni assonometrici CAD allegati alla documentazione di scavo, durante la nuova attività di ricerca e analisi dei dati archeologici, condotta dal 2018[13], è stato sviluppato un modello tridimensionale adoperando i software commerciali SketchUp Pro 2021[14] e OneRay-RT[15], con lo scopo di comprendere, visualizzandolo, l'ultimo assetto del monastero (Fig. 2). Nonostante i limiti di tale modello, relativi ancora alla geometria schematica degli elevati ricostruiti e alle *texture* ottenute da rilievi fotogrammetrici delle poche murature superstiti, la ricostruzione 3D ha consentito di comprendere i volumi del monastero nel periodo della sua massima espansione.

Figura 2. *Ricostruzione 3D del monastero Di Sant'Ilario a Port'Aurea (S. Rapuano)*

[12] Le indagini sono state realizzate dall'allora Soprintendenza Archeologica di Salerno, Avellino e Benevento, in collaborazione con la Provincia e il Comune di Benevento, sotto la direzione scientifica di Giuseppina Bisogno.

[13] Grazie all'autorizzazione concessa dall'attuale Soprintendenza Archeologia, Belle Arti e Paesaggio per le province di Caserta e Benevento a Marcello Rotili e alla sua equipe, è stato condotto uno studio sistematico sui dati di scavo, rimasti a lungo inediti e privi di un adeguato approfondimento, e sui numerosi reperti, di cui sono stati realizzati schedatura e rilievi (Rapuano 2022).

[14] https://www.sketchup.com/.

[15] https://www.analistgroup.com/it/software-rendering-3D.

MODUS

Il progetto di virtualizzazione della chiesa di Santa Sofia di Benevento, gemmato da un filone di studi pluriennale, è stato avviato nel gennaio 2020, al fine di verificare una serie di ipotesi archeologiche. A partire dalla raccolta della documentazione storica-archeologica e dal rilievo e dalla modellazione 3D della chiesa nella sua configurazione attuale[16], sono stati ricostruiti i principali assetti planimetrici che l'edificio ha assunto nel tempo a causa di numerosi restauri e rimaneggiamenti[17] (Fig. 3-5).

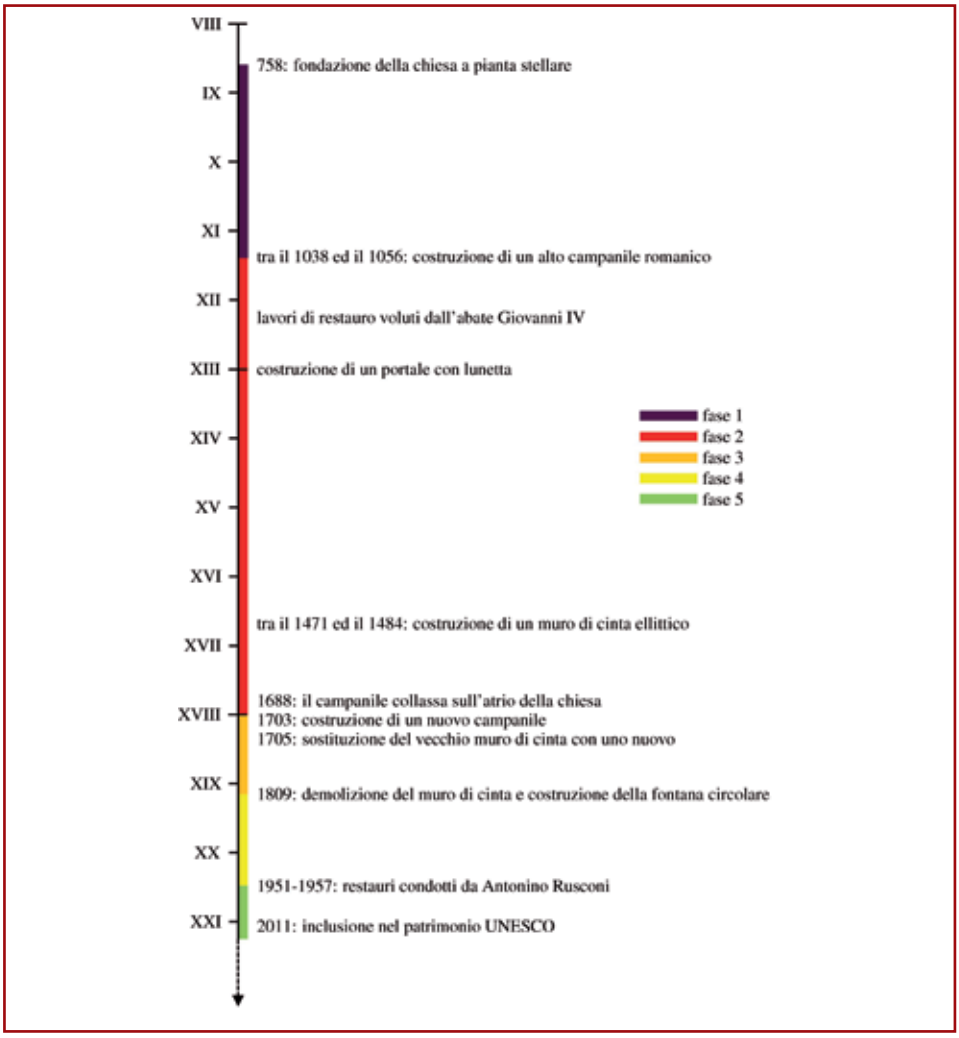

Figura 3. *I principali eventi noti della storia di Santa Sofia e indicazioni delle fasi storico-archeologiche (Babilio – Rapuano 2025b, fig. 2)*

[16] La chiesa, a pianta centrale, presenta un muro d'ambito parzialmente a zig-zag e, nella zona del presbiterio, tre absidi circolari contigue. Lo spazio interno, scandito da otto colonne, composte da *spolia* di età romana, e otto pilastri a base quadrata, è coperto da un sistema di volte di varia forma e una cupola al centro.

[17] Babilio — Rapuano 2023; 2024; 2025a-b.

Fu fondata nel 758 d.C. da Arechi II, duca di Benevento, e dedicata alla Divina Sapienza di Cristo, ispirandosi all'omonima chiesa di Santa Sofia di Costantinopoli (fase 1). In età romanica fu aggiunto un campanile e realizzato il portale con lunetta (fase 2). Quest'ultimo è tutt'oggi visibile nella facciata. Gravemente danneggiata dai terremoti del 1688 e del 1702 (che provocarono anche il crollo del campanile romanico), fu restaurata dall'arcivescovo Vincenzo Maria Orsini, che trasformò la pianta stellata originale in una forma circolare e aggiunse una cappella centrale rettangolare, demolendo l'abside centrale[18] (fase 3). Ulteriori interventi nel XIX secolo portarono alla costruzione di due cappelle laterali e alla demolizione del muro che chiudeva lo spazio antistante la chiesa (fase 4). Tra il 1951 e il 1957, il Soprintendente ai Monumenti, Antonino Rusconi, ripristinò la pianta stellata originale, restaurò le absidi laterali e ricostruì quella centrale preservando la facciata barocca[19]. Dal 2011, la chiesa è parte del sito UNESCO "I Longobardi in Italia: i luoghi del potere" (fase 5).

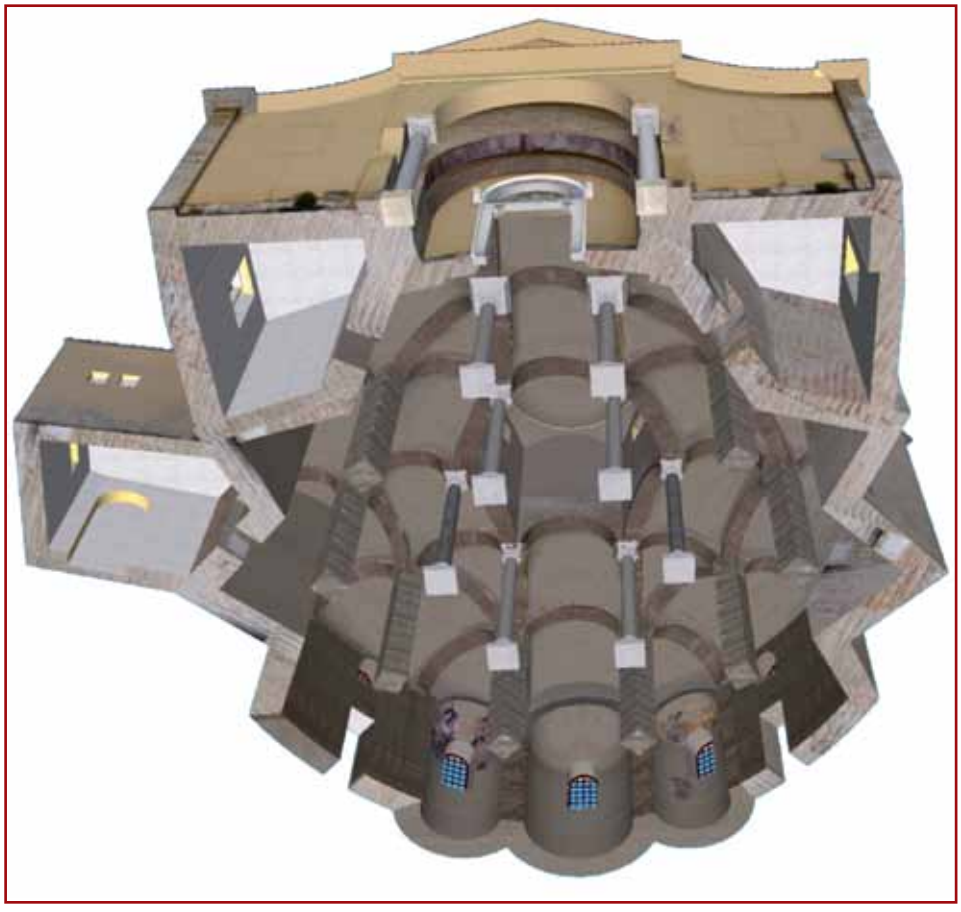

Figura 4. *Ricostruzione 3D della chiesa di Santa Sofia, assetto attuale (E. Babilio, S. Rapuano)*

[18] Rotili 2017; Rusconi 1967.
[19] Rotili — Pace 2021, p. 81-82.

MODUS

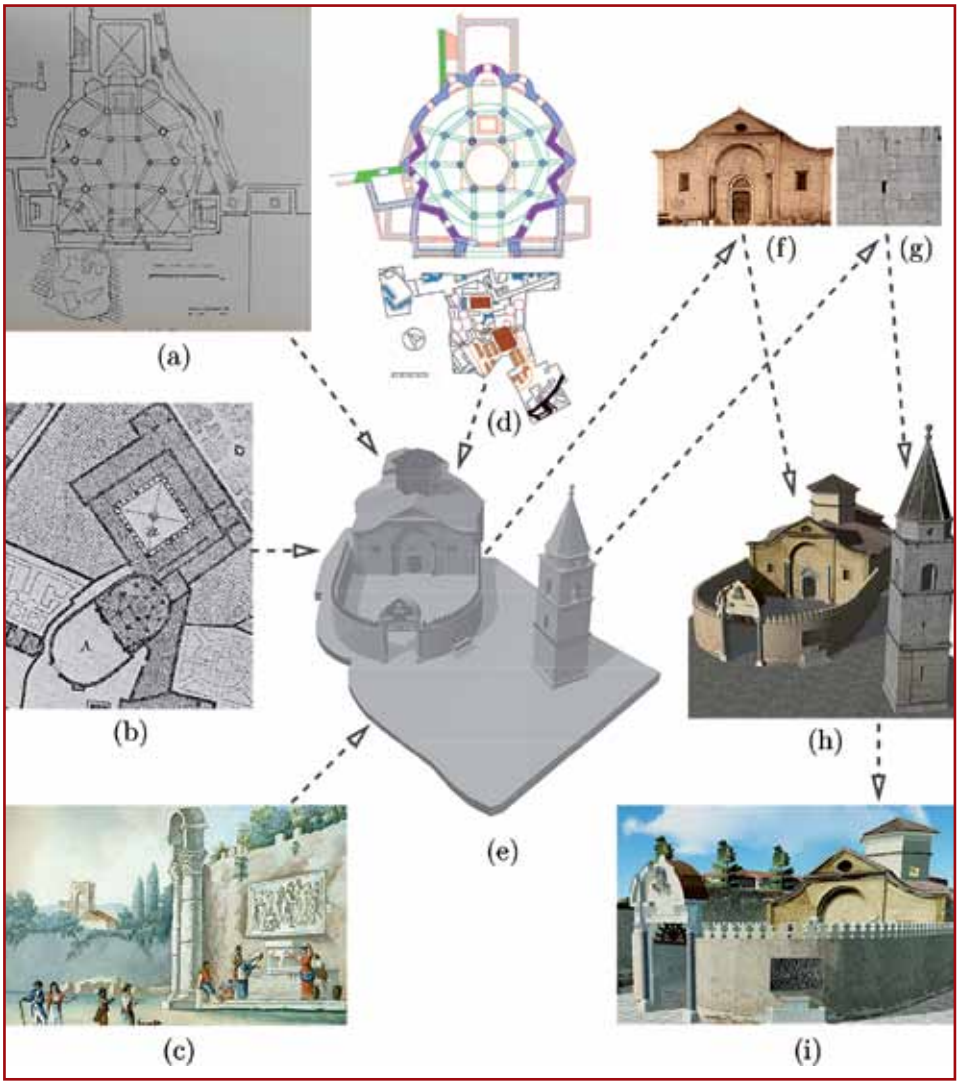

Figura 5. *Uno schema della virtualizzazione della terza fase storica-archeologica di Santa Sofia, a partire dalla raccolta dei dati: (a) disegni tecnici; (b) dettagli tratti da mappe storiche della città di Benevento; (c) dipinti; (d) informazioni provenienti da scavi archeologici; (e) il modello solido; texture di (f) intonaco della chiesa e (g) superficie del campanile; (h) modello texturizzato; (i) modello renderizzato (Babilio – Rapuano 2024, fig. 2)*

Per lo studio sono stati utilizzati: il laser scanner Leica BLK360[20] e Cyclone Register per il rilievo e la gestione della nuvola di punti; Autodesk AutoCAD 2020 per la modellazione solida; SketchUp per realizzazione di *texture* da foto e per la loro applicazione al modello solido; OneRay-RT per il rendering fotorealistico.

[20] Leica Geosystems AG 2017.

Inoltre, per la virtualizzazione del complesso sistema di volte, è stato necessario adottare uno specifico *workflow*:

– estrazione delle volte dalla nuvola di punti con 3DF Zephyr Free;
– esportazione delle coordinate in un file ASCII;
– realizzazione di uno *script* (*notebook*) *ad hoc* in ambiente Mathematica®[21] per la scrittura simbolica di funzioni polinomiali per approssimare le superfici delle volte;
– esportazione delle superfici in un file DXF e integrazione nel modello 3D in AutoCAD.

In aggiunta, per rendere i risultati dello studio accessibili anche ad un pubblico di non esperti e favorire la conoscenza e la valorizzazione del sito, è stato realizzato con il software Animotica un filmato per la piattaforma YouTube[22] (Fig. 6).

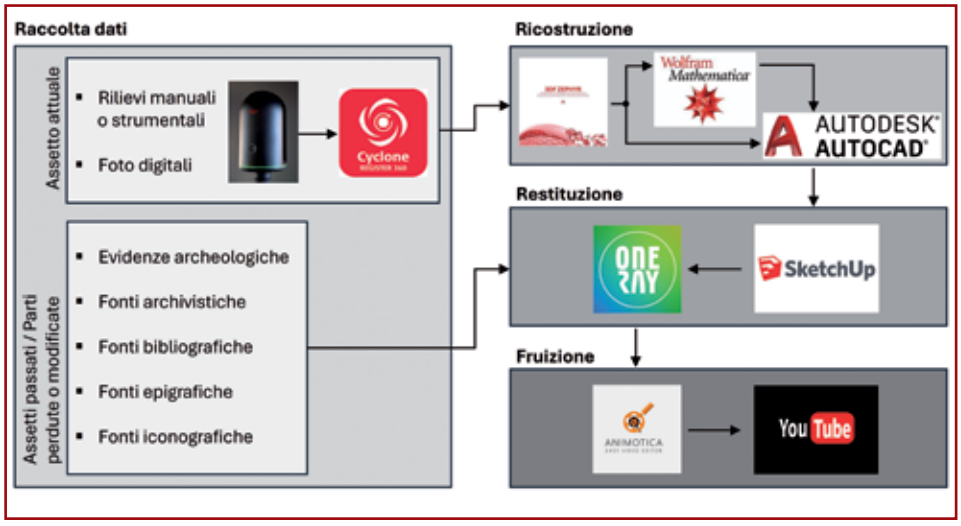

Figura 6. *Workflow per la virtualizzazione della chiesa di Santa Sofia (E. Babilio, S. Rapuano)*

3. Nuovi approcci informatici nella ricostruzione archeologica

Con la breve illustrazione dei modelli appena esposti si è inteso riassumere modalità di esecuzione e procedimenti specifici basati sulle esperienze, sulle finalità prefisse e sulla disponibilità di risorse del team di ricerca, che dunque hanno una valenza particolare. Al contrario, per quanto riguarda le tecnologie impiegate, esse corrispondono a quelle più comunemente adottate in ambito archeologico. A sostegno di questa osservazione appare utile e interessante consultare i dati pubblicati a seguito di un sondaggio internazionale, realizzato tra gennaio e marzo 2020 dalla *European Association of Archaeologists* (EAA) *Community for the use and application of 3D technologies in Archaeology*

[21] Wolfram Research, Inc.: Mathematica, Version 8.0, champaign, IL (2010).
[22] https://www.youtube.com/watch?v=KmpUnn2QxcE&t=27s.

(*3D-Archaeology*). L'obiettivo perseguito da tale operazione è stato quello di acquisire una conoscenza più ampia dell'applicazione delle tecnologie 3D nella pratica archeologica e nella gestione del patrimonio culturale, ma soprattutto di riflettere sui possibili sviluppi e implementazioni.
Il sondaggio ha permesso di fornire una panoramica sia dei software e dei formati di dati utilizzati, sia delle attuali pratiche di archiviazione dei dati grezzi e/o generati.
Il programma Metashape di Agisoft LLC è risultato di gran lunga il più utilizzato, seguito da ReCap Pro di Autodesk e RealityCapture di Capturing Reality. Meshroom di AliceVision è stato l'unico software open source menzionato tra quelli più usati.
Le applicazioni utilizzate possono essere, per la maggior parte, classificate come "soluzioni complete", che includono anche strumenti per la georeferenziazione.
Dal sondaggio è emerso che i formati di dati per l'input più usati sono JPEG, TIFF e RAW e per l'output sono OBJ, PLY e GeoTIFF.
Inoltre, è risultato che coloro che operano in questo settore sono liberi professionisti, istituti di ricerca e autorità preposte alla tutela del patrimonio.
Il sondaggio ha anche evidenziato che è necessario monitorare con maggiore attenzione tali prodotti e migliorarne la standardizzazione per favorire la comunicazione, con l'auspicio di un coinvolgimento maggiore di archeologi e di altri professionisti del patrimonio culturale e di settori correlati[23].

Proprio sulla standardizzazione e la trasparenza dei prodotti 3D la comunità scientifica si sta interrogando. La diffusione dei modelli ottenuti deve essere accompagnata da informazioni dettagliate sull'oggetto originale e sulle fasi di conversione in risorsa digitale. È fondamentale registrare dati come calibrazione, tolleranza, errore strumentale e condizioni fisiche al momento dell'acquisizione, oltre alle fasi di post-elaborazione. Questi dettagli garantiscono trasparenza e permettono una corretta interpretazione e riutilizzo del modello. Diversi progetti europei, tra i quali si citano CARARE, 3D-ICONS, 3D-COFORM, ARIADNE, PARTHENOS, hanno sviluppato sistemi di metadati specifici per il 3D, che documentano tecniche e modalità di acquisizione, migliorando la condivisione e la conservazione a lungo termine delle risorse digitali. Dunque, associare ai modelli digitali informazioni complete è un'attività necessaria per evitare che la semplificazione delle tecniche di acquisizione porti a repliche prive di valore scientifico[24]. Seguendo tali criteri, la realizzazione dei modelli 3D diacronici dei siti beneventani, è stata condotta mediante la consultazione di fonti attendibili, la rappresentazione grafica dei dati archeologici, la comunicazione visiva delle ipotesi ricostruttive del sito e la disseminazione dei procedimenti e dei metadati attraverso pubblicazioni scientifiche, ossia gli stessi principi che sottendono a progetti di più ampio rilievo, come, ad esempio, l'Extended Matrix, un progetto open science che consente di creare, visualizzare, esplorare e condividere

[23] Opgenhaffen — Jeffra — Hilditch 2024, p. 131-150.
[24] Bosco — D'Andrea 2019.

online modelli 3D utilizzando piattaforme gratuite e open source[25]. Sviluppato e testato dal laboratorio Virtual Heritage dell'Istituto di scienze del patrimonio culturale del Consiglio nazionale delle ricerche (ISPC CNR), si ispira al metodo stratigrafico della *Matrix* di Harris. Si basa sulla metodologia specifica dell'HBIM (Heritage Building Information Model, evoluzione della già citata BIM) applicata alla ricostruzione e analisi dei contesti archeologici post-scavo e documentazione. Questo approccio permette di integrare e coordinare vari tipi di dati: dai reperti archeologici ai rilievi, dalle evidenze stratigrafiche alle fonti bibliografiche, sino ai confronti. Tutto ciò serve a formulare ipotesi ricostruttive che possano essere sostenute da livelli di affidabilità differenziati. Collegato a Extended Matrix è il *framework* open source ATON, sviluppato sempre dal laboratorio Virtual Heritage del ISPC CNR, rappresenta una soluzione avanzata per la progettazione di applicazioni web immersive e collaborative nel settore del patrimonio culturale. Questo strumento è concepito per facilitare l'accesso a contenuti 3D interattivi attraverso molteplici dispositivi, sfruttando appieno le tecnologie web e i recenti standard del settore. Grazie alla sua architettura, consente di usufruire di applicazioni tridimensionali interattive direttamente tramite un comune browser, eliminando così la necessità di installazioni da parte dell'utente finale[26].

Un ulteriore sviluppo nei progetti di archeologia virtuale è l'applicazione dell'*Artificial Intelligence* (AI). Può essere usata per analisi statistiche complesse su grandi volumi di dati, per lo sviluppo di script GIS, per l'identificazione automatica di reperti archeologici, sino alla creazione di ricostruzioni grafiche realistiche di siti e manufatti. Un esempio di progetto basato su sistema integrato di AI, sensoristica archeometrica e robotica è AUTOMated enriched digitisation of Archaeological liThics and cerAmics (AUTOMATA), sviluppato dall'Università di Pisa, per facilitare lo studio dei reperti. Consiste in un braccio robotico che preleva i frammenti ceramici, li analizza con sensori per il riconoscimento delle proprietà chimico-fisiche e infine genera un modello 3D completo di tutti dati. Il lavoro di documentazione di AUTOMATA è destinato ad essere riversato nel cloud dedicato alla condivisione e all'utilizzo innovativo del patrimonio culturale che l'Unione Europea sta costruendo (ECCCH: European Collaborative Cloud for Cultural Heritage - ECHOES project: European Cloud for Heritage OpEn Science)[27].

[25] Demetrescu — Ferdani 2021.
[26] Fanini — Ferdani — Demetrescu — Berto — d'Annibale 2021.
[27] <https://old.unipi.it/index.php/news/item/28902-automata-l-archeologia-diventa-robotica-e-smart>; <https://www.cfs.unipi.it/progetti/automata-automated-enriched-digitisation-of-archaeological-lithics-and-ceramics/>.

4. Considerazioni su alcuni limiti della modellazione solida applicata all'archeologia

Come accennato nei paragrafi precedenti, esiste una grande varietà di soluzioni per la gestione dei dati di rilievo tridimensionale e la modellazione solida e la scelta degli strumenti può dipendere da molti fattori. Tuttavia, quasi mai un unico strumento è sufficiente. Anzi, generalmente, diviene obbligatorio integrarne diversi. Questo aspetto rappresenta ancora un limite dell'attività di realizzazione di modelli 3D e gemelli virtuali, per la insita necessità di acquisire *know-how* specifici. Un'ulteriore questione di rilevante interesse è certamente legata ai costi, spesso elevati che la tecnologia utilizzata, sia hardware che software, presenta, sebbene questa voce sia soggetta a variazione dovute al mercato. Per quanto riguarda i costi del software, quelli commerciali tendono a restare generalmente invariati da una versione all'altra di uno specifico programma, anche se formule di abbonamento possono consentire qualche abbattimento. In ogni caso, l'esistenza di software freeware o addirittura open source può consentire addirittura l'azzeramento dei costi per il software, sebbene questo tipo di soluzioni in molti casi presenti una curva di apprendimento più ripida delle soluzioni commerciali. Infine, da un punto di vista puramente scientifico, permane la necessità di protocolli codificati, chiari e sempre verificabili che garantiscano che un prodotto digitale, in particolare per la ricerca archeologica, abbia caratteristiche di affidabilità e riproducibilità, al pari di ogni altra tipologia di esperimento scientifico.

5. Note conclusive

L'impiego dei modelli digitali 3D ha profondamente cambiato il panorama della ricerca archeologica, portando a una valorizzazione senza precedenti del patrimonio culturale. Attraverso esempi concreti, come la virtualizzazione dei siti medievali di Benevento, è emerso come queste tecnologie non solo facilitino l'interazione con i beni culturali ma siano anche strumenti chiave per la verifica di ipotesi storiche e architettoniche. Tuttavia, è essenziale riconoscere i limiti intrinseci a queste pratiche, quali la necessità di competenze specializzate e le barriere economiche legate all'uso degli strumenti. L'adozione di protocolli chiari e standardizzati risulta cruciale per garantire l'affidabilità e la validità scientifica dei modelli 3D.

L'integrazione di tecnologie avanzate e il miglioramento della condivisione delle risorse digitali offriranno nuove opportunità, ma richiederanno anche un impegno collaborativo.

Attualmente, il processo che consente di duplicare un bene fisico d'interesse archeologico è difficilmente automatizzabile nella sua interezza.

In una prospettiva futura ci si può attendere che questo limite operativo sia superato con un utilizzo sempre più massiccio di algoritmi di intelligenza artificiale. Questi, grazie ad automatizzazione e standardizzazione, se da un lato consentiranno la diminuzione della mole di lavoro richiesto, dall'altro presenteranno il concreto rischio di perdere il controllo sul flusso dei dati. Dunque, quanto maggiore sarà il ricorso ad algoritmi, tanto più

specializzato dovrà essere l'approccio critico-disciplinare al fine di garantire affidabilità e accuratezza di processi logici e riproducibilità dei risultati. In quest'ottica, il contributo dell'archeologo rimane imprescindibile.

The aim of this contribution is to consider the development and use of three-dimensional (3D) digital models in archaeological research. Realising a 3D model of an existing (or planned) artefact makes it possible to manage and combine so much information that the model can be considered the virtual counterpart, the so-called digital twin, of the represented physical system. In archaeology, through digital twins, it is possible to interact and study sites or artefacts and, for research purposes, verify hypotheses about their historical and architectural phases. Furthermore, a 3D model can be a valuable tool for the dissemination of heritage. These issues will be explored starting from a series of experiences of virtual reconstructions of medieval sites in the city of Benevento, an urban reality characterised by multiple transformations and settlement events. The procedure that led to the implementation of these products, the results obtained in terms of understanding and analysing the sites and the prospects for the valorisation and dissemination of scientific data will be illustrated, also considering the limits and criticalities of this type of research.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

Riferimenti bibliografici

Babilio — Rapuano 2023

Enrico Babilio — Silvana Rapuano. *The Complex Architecture of the Vault System of an Early Medieval Church*. «*Heritage*», 6 (2023), p. 5779-5804.
<https://doi.org/10.3390/heritage6080304>.

Babilio — Rapuano 2024

Enrico Babilio — Silvana Rapuano. *Digital Twin Model of an Early Medieval Church: Entanglement of Historical Studies and Mathematical Methods*. In: *9th European Congress on Computational Methods in Applied Sciences and Engineering* (ECCOMAS). DOI: 10.23967/eccomas.2024.093.
<https://www.scipedia.com/public/Babilio_Rapuano_2024a>.

Babilio — Rapuano 2025a

Enrico Babilio — Silvana Rapuano. *Virtualization for Knowledge and Protection: the Case of an Early Medieval Church*. In: *Protection of Historical Constructions. PROHITECH 2025. Lecture Notes in Civil Engineering*, a cura di F. M. Mazzolani, R. Landolfo, B. Faggiano. Cham: Springer International Publishing, 2025 (v. 596).
<https://doi.org/10.1007/978-3-031-87316-4_59>.

Babilio — Rapuano 2025b

Enrico Babilio — Silvana Rapuano. *The Main Phases of an Early Medieval Church: Virtual Reconstructions for Cultural Heritage Promotion*. In: *The Future of Heritage Science and Technologies II, Design, Simulation and Monitoring*, a cura di R. Furferi, L. Governi, Y. Volpe, F. Gherardini, K. Seymour. Cham: Springer International Publishing, 2025.
<https://doi.org/10.1007/978-3-031-98379-5_38>.

Bosco — D'Andrea 2019

Angela Bosco — Andrea D'andrea. *Documentazione grafica per lo studio dei beni archeologici attraverso la replica digitale 3D*. In: *Riflessioni L'Arte del Disegno / Il Disegno dell'Arte, Atti del 41° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione Congresso della Unione Italiana per il Disegno - UID2019*, Perugia, 19-20-21 Settembre 2019, a cura di P. Beraldi. Roma: Gangemi Editore, 2019, p. 425-430.

Demetrescu — Ferdani 2021

Emanuel Demetrescu — Daniele Ferdani. *From Field Archaeology to Virtual Reconstruction: a Five Steps Method Using the Extended Matrix*. «Applied Sciences», 11 (2021), 5206.
<https://doi.org/10.3390/app11115206>.

Fanini — Ferdani — Demetrescu — Berto — d'Annibale 2021.

Bruno Fanini — Daniele Ferdani — Emanuel Demetrescu — Simone Berto — Enzo d'Annibale. *ATON: an Open-Source Framework for Creating Immersive, Collaborative and Liquid Web-Apps for Cultural Heritage*. «Applied Sciences», 11 (2021), 11062.
<https://doi.org/10.3390/app112211062>.

Gabellone 2019 — Francesco Gabellone. *Archeologia Virtuale: teoria, tecniche e casi di studio*. Lecce: Edizioni Grifo, 2019.

Hutson — Weber — Russo 2023 — James Hutson — Joseph J. Weber — Angela Russo. *Digital twins and cultural heritage preservation: a case study of best practices and reproducibility in chiesa dei SS Apostoli e Biagio*. «Art and Design Review», 11 (Febbraio 2023), n. 1, p. 15-41. <https://doi.org/10.4236/adr.2023.111003>.

Leica Geosystems AG 2017 — *Leica BLK360 User Manual*. Heerbrugg (Switzerland), Version 2.0 (2017), 853811-2.0.0en.

Liu — Lu — Chen — Yan 2024 — Wanqin Liu — Man Lu — Yuqin Chen — Kaikai Yan. *Digital twin applications in an archaeological site: a virtual reconstruction of the Pishan site, Zhejiang, China*. «Journal of Chinese Architecture and Urbanism», 6 (2024), n. 1, 1735. <https://doi.org/10.36922/jcau.1735>.

Niccolucci — Felicetti 2024 — Achille Felicetti — Franco Niccolucci. *Il futuro digitale di 4CH: il gemello digitale dei beni culturali.* «DigItalia. Rivista del digitale nei beni culturali», 19 (2024), n. 2, p. 117-125. <https://doi.org/10.36181/digitalia-00105>.

Opgenhaffen — Jeffra — Hilditch 2024 — Loes Opgenhaffen — Caroline Jeffra — Jill Hilditch. *Balancing Data Storage and User Functionality: the 3D and Archaeological Data Strategy of the Tracing the Potter's Wheel Knowledge Hub.* In: *The 3 Dimensions of Digitalised Archaeology State-of-the-Art, Data Management and Current Challenges in Archaeological 3D-Documentation*, a cura di M. Hostettler, A. Buhlke, C. Drummer, L. Emmenegger, J. Reich, C. Stäheli. Cham: Springer International Publishing, 2024, p. 131-150. <https://doi.org/10.1007/978-3-031-53032-6_1>.

Parry 2005 — Ross Parry. *Digital heritage and the rise of theory in museum computing*. «Museum Management and Curatorship», 20 (2005), n. 4, p. 333-348. DOI:10.1016/j.musmancur.2005.06.003.

Pasini 2004 — Giovanna Pasini. *Tecnica digitale. Una nuova estetica per i Beni Culturali.* «La Rivista di Engramma», 33, (May 2004), p. 19-24. DOI: 10.25432/1826-901X/2004.33.0007.

Pfarr-Harfst — Grellert 2016 — Mieke Pfarr-Harfst — Marc Grellert. *The reconstruction – argumentation method.* In: *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, a cura di M. Ioannides, E. Fink, A. Moropoulou, M. Hagedorn-Saupe, A. Fresa, G. Liestøl, V. Rajcic, P. Grussenmeyer. Cham: Springer International Publishing, 2016, p. 39–49.

MODUS

Rapuano 2022 — Silvana Rapuano. *Archeologia e Storia di un monastero. Sant'Ilario a Port'Aurea di Benevento*. Bari: Edipuglia, 2022 (Archeologia Postclassica; 13).

Rapuano — Busino — Rotili 2016 — Silvana Rapuano — Nicola Busino — Marcello Rotili. *Dalla diagnostica alla valorizzazione in un contesto romano-medievale: il caso di Benevento*. In: *Stratigrafia degli elevati e nuove tecnologie diagnostiche. Archeologia dell'edilizia storica in situazioni d'emergenza*, a cura di F. Redi, A. Forgione, F. Armillotta. L'Aquila: One Group Edizioni, 2016, p. 63-80.

Rotili 2006 — Marcello Rotili. *Cellarulo e Benevento, la formazione della città tardoantica*. In: *Benevento nella Tarda Antichità. Dalla diagnostica archeologica in contrada Cellarulo alla ricostruzione dell'assetto urbano*, a cura di M. Rotili. Napoli: Arte tipografica, 2006, p. 9-88.

Rotili 2017 — Marcello Rotili. *Arechi II e Benevento*. In: *Tra i Longobardi del Sud: Arechi II e il Ducato di Benevento,* a cura di M. Rotili. Padova: Il poligrafo, (2017), p. 181-226.

Rotili — Pace 2021 — Marcello Rotili — Valentino Pace. *Benevento, chiesa di Santa Sofia*. In: *Studi sulla Pittura Beneventana. II. Aggiornamento Scientifico Hans Belting*, a cura di G. Bertelli, M. Mignozzi. Bari: Mario Adda Editore, 2021, p. 79-93.

Rotili — Rapuano — Cataldo 2010 — Marcello Rotili — Silvana Rapuano — Maria Raffaella Cataldo. *Nuovi dati su Benevento nella tarda antichità*. In: *Paesaggi e insediamenti urbani in Italia meridionale fra tardoantico e alto medioevo (STAIM 2), Atti del secondo seminario sul tardoantico e l'alto medioevo in Italia meridionale (Foggia-Monte Sant'Angelo, 27-28 maggio 2006)*, a cura di G. Volpe, R. Giuliani. Bari: Edipuglia, 2010, p. 309-328.

Rusconi 1967 — Antonino Rusconi. *La chiesa di Santa Sofia di Benevento*. In: *XIV Corso di Cultura sull'Arte Ravennate e Bizantina*. Ravenna: Longo Editore, 1967, p. 339–359.

MODUS

Progetti

Fondi, Possessori ed Esemplari nell'Indice SBN

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00143

Maria Cristina Mataloni — Alice Semboloni
Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU)

Nell'ambito del progetto Indice 3 relativo alla reingegnerizzazione ed evoluzione dell'Indice del Servizio Bibliotecario Nazionale (SBN), l'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU) ha previsto la possibilità di gestire a livello centrale i fondi di persona/ente presenti nelle biblioteche SBN e finora gestiti solo a livello locale. In concreto, l'Istituto sta sviluppando il progetto Fondi-Possessori-Esemplari che mette in relazione tra di loro questi elementi. La centralizzazione in Indice SBN delle informazioni relative all'esemplare, al fondo, a provenienza/possessore e ai legami reciproci permetterà una gestione condivisa con tutte le biblioteche SBN per mezzo di un ampliamento dell'attuale struttura. Ad oggi, ciascun polo SBN che tratta questa tipologia di dati, lo fa in completa autonomia generando un proliferare di sistemi dissimili e offrendo ai propri utenti una possibilità di interrogazione molto diversificata e parziale. Con questo nuovo progetto, di rilevanza non solo per la comunità SBN, gli utenti avranno un punto di accesso unico per la ricerca integrata sui fondi, i possessori e gli esemplari.

Introduzione

Negli ultimi anni, nell'ambito degli studi bibliografici, si è riscontrato un crescente interesse per aspetti legati alla circolazione del libro, alla sua distribuzione e diffusione, nonché al suo pubblico. Ne sono prova i numerosi convegni organizzati in Italia negli ultimi anni, seguiti da un pubblico sempre maggiore di studiosi e appassionati del settore, e importanti progetti e iniziative nazionali che hanno avuto come oggetto lo studio degli esemplari - e quindi dei "segni" della propria storia che reca la singola copia - dei fondi conservati dalle biblioteche e dei possessori[1], anche con la realizzazione di archivi controllati (di persone, famiglie e istituzioni)[2].

[1] Per avere un'idea della quantità dei progetti, si rinvia al dossier relativo agli archivi controllati di possessori: *Cataloghi, biblioteche e dati di esemplare: un dossier internazionale sulle banche dati delle provenienze = Catalogues, libraries and copy-specific evidence: an international dossier on provenance databases*, «La Bibliofilía», 117 (2015), n. 3, p. 309-366, e in particolare al contributo in esso pubblicato di Luca Rivali dal titolo: *Storia del libro e provenienze: introduzione al dossier*, p. 309-317. In merito alle iniziative nazionali nell'ambito del libro moderno e contemporaneo, e in particolare delle biblioteche d'autore, si veda l'attività della Commissione nazionale biblioteche speciali, archivi e biblioteche d'autore dell'AIB, <https://www.aib.it/struttura/commissioni-e-gruppi/gbaut/>, che ha pubblicato le linee guida sul trattamento dei fondi personali: <https://www.aib.it/struttura/commissioni-e-gruppi/gbaut/strumenti-di-lavoro/linee-guida-sul-trattamento-dei-fondipersonali/>.

[2] Per una panoramica sulle principali basi dati si rimanda all'articolo di Flavia Bruni, *Per un indice condiviso di possessori e provenienze in SBN: una prospettiva concreta*, «AIB studi», 60 (2020), n. 2, p. 293-309, <https://doi.org/10.2426/aibstudi-12262>.

MODUS

L'importanza storica della ricostruzione bibliografica delle raccolte è stata ribadita più volte dagli studiosi. Lorenzo Baldacchini scrive che:

> «la funzione documentaria accanto all'aspetto che potremmo definire "archeologico" che riguarda il libro come singolo manufatto, ne comprende un altro, bibliografico, da qualcuno paragonato impropriamente a quello archivistico, che riguarda il libro inserito in un insieme più vasto (raccolta, fondo, biblioteca), e quindi i suoi rapporti con altri libri: appartenenza ad un fondo, provenienza di questo e sua eventuale fusione con altri, sistemazione e organizzazione del fondo e del singolo pezzo al suo interno»; «nonostante molti si ostinino ancora a negarlo, esiste un preciso rapporto con il territorio [...] anche dei beni librari, considerati non come singolo e magari "raro" pezzo, ma come insieme di raccolte che sono ad un tempo espressione e funzione dello sviluppo culturale di una realtà, regionale, cittadina, al limite di quartiere».

E aggiunge: «dovremmo perciò tenere conto di questo duplice aspetto del libro antico per trattarlo correttamente»[3].

Da un punto di vista pratico, Alberto Petrucciani ribadisce «che il modo di procedere più opportuno sia oggi, non quello di censimenti basati su formali criteri bibliografici (di cinquecentine, di seicentine, ecc.), ma quello di una ricognizione integrale, topografica e sistematica, delle collezioni»[4]; e sottolinea l'importanza di un "catalogo integrato" di tutti i materiali – che costituiscono un insieme unitario.

Sulla base di queste esigenze, l'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU)[5] ha deciso di dare vita al progetto Fondi-Possessori-Esemplari con la finalità di offrire uno strumento aggiuntivo nel Servizio Bibliotecario Nazionale (SBN)[6], di rilevanza nazionale per lo studio dei fondi. Il progetto si basa sulla condivisione a livello centrale, nell'Indice SBN, delle informazioni relative agli esemplari e ai loro legami con provenienze[7] e possessori, che potranno essere arricchite da ulteriori informazioni descrittive del fondo a cui appartengono grazie all'integrazione con la base dati dell'Anagrafe delle biblioteche italiane.

Ad oggi, non esiste infatti a livello nazionale una struttura unica che raccolga questi dati e ne permetta l'interrogazione. La gestione dei dati relativi a fondi, possessori/prove-

[3] Lorenzo Baldacchini, *Il libro antico*, Roma: Carocci, 2001, p. 16.

[4] Alberto Petrucciani, *Dai censimenti bibliografici alla storia della cultura e della società: riflessioni sul ruolo delle biblioteche tra ricerca e comunità*, «Nuovi annali della scuola speciale per archivisti e bibliotecari», 34 (2020), p. 257-270.

[5] https://www.iccu.sbn.it/it/.

[6] SBN è la più importante infrastruttura culturale italiana. Ha una struttura stellare in cui al centro si trova l'Indice SBN, che è collegato ai poli SBN. Questi ultimi, a loro volta, permettono alle biblioteche aderenti a ciascun polo di scambiare con l'Indice le informazioni sulle schede catalografiche e sugli elementi di Authority (Nomi, Soggetti, Luoghi, Marche, Titoli dell'Opera). Per ulteriori approfondimenti: <https://www.iccu.sbn.it/it/SBN/>.

[7] Si definisce "provenienza" un particolare possessore, ovvero quello che ha preceduto l'attuale. Cfr. l'introduzione di Luca Rivali al dossier sulle banche dati di provenienza citato alla nota 1. La distinzione terminologica convenzionale tra possessore e provenienza è stata recepita e adottata nell'ambito del Servizio Bibliotecario Nazionale, si veda: Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, *Guida alla catalogazione in SBN: materiale antico*, a cura dell'area Attività per la bibliografia, la catalogazione e il censimento del libro antico, Roma: ICCU, 2016, <https://norme.iccu.sbn.it/images/8/8e/Guida_SBN _Antico.pdf>.

MODUS

nienza - esemplari, in questi anni, e in carenza di un progetto nazionale, è stata portata avanti a livello locale da alcuni poli e biblioteche, con iniziative meritevoli, ma limitate necessariamente a una possibilità di interrogazione parziale e frammentaria.
La creazione di una struttura condivisa consentirà alle biblioteche SBN di inviare a livello centrale questa tipologia di dati, che ad oggi sono ospitati in sistemi dissimili, sviluppati in autonomia dagli istituti. Gli studiosi potranno beneficiare in questo modo di un punto di accesso unico per la ricerca, sfruttando l'enorme ricchezza dell'Indice SBN per scoprire inedite connessioni e relazioni tra i documenti.
Il progetto è alle fasi conclusive di realizzazione per quanto riguarda gli aspetti di back-end, consistenti nella creazione di una infrastruttura per accogliere i dati in Indice SBN. Seguirà, a questa fase di sviluppo, la possibilità di ricerca con apposita funzione nell'Opac SBN[8], con la realizzazione degli indici e dell'interfaccia utente.
La struttura predisposta in Indice verrà progressivamente popolata, grazie all'apporto delle biblioteche, e sarà messa a disposizione dell'utenza nell'Opac SBN non appena il numero dei dati consentirà di effettuare un'efficace ricerca.
Il progetto si inserisce all'interno del più ampio disegno di reingegnerizzazione dell'Indice SBN, il progetto Indice 3 che, oltre all'aggiornamento tecnologico, prevede diversi interventi evolutivi.

Perché un progetto per Fondi-Possessori-Esemplari

Nell'Indice SBN non esistono al momento dati pertinenti alle diverse copie della medesima notizia bibliografica presenti nella stessa o in diverse biblioteche. Ci riferiamo, ad esempio, ai dati gestionali come inventario e collocazione. Questa impostazione risponde alla filosofia e all'architettura di SBN, in cui all'Indice è assegnato il compito di registrare solamente le informazioni condivisibili tra tutte le biblioteche della rete. Ciò consente, tra l'altro, un traffico limitato nello scambio di dati Polo-Indice. Informazioni quali i dati inventariali e di collocazione, la disponibilità dei materiali, le informazioni relative all'utenza bibliotecaria, sono presenti solo nelle basi dati periferiche dei poli SBN (Fig. 1).

[8] https://opac.sbn.it/.

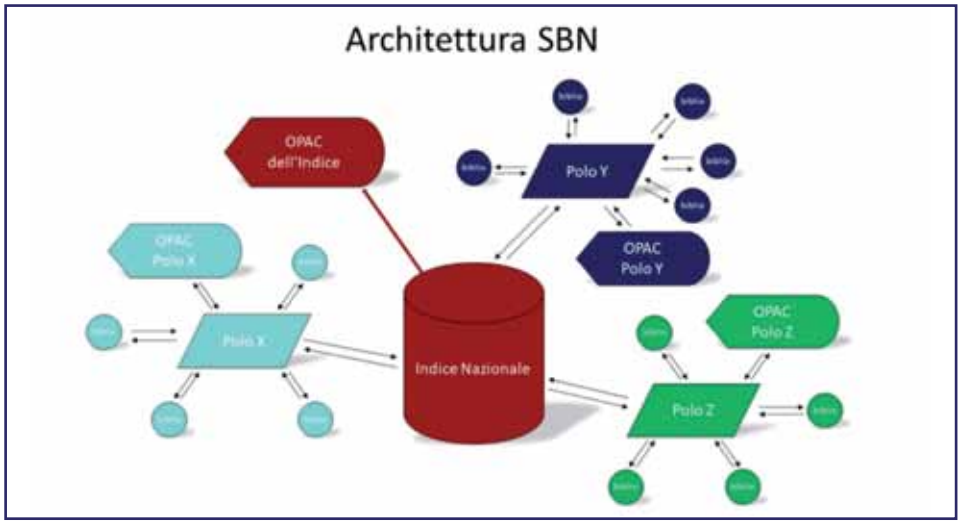

Figura 1. *Architettura di SBN: al centro si trova l'Indice SBN che colloquia con le basi dati dei poli e condivide con essi parte dei dati. I poli SBN sono strutture tecnico-amministrative che permettono alle biblioteche aderenti al polo di scambiare informazioni con l'Indice. Nelle basi dati di polo si trova una tipologia di dati maggiore rispetto a quelli presenti in Indice*

I dati condivisi sono quelli visibili nell'Opac SBN e permettono l'individuazione del documento di interesse e delle biblioteche che lo possiedono. Attraverso le localizzazioni dell'Opac SBN l'utente può accedere direttamente ai cataloghi locali che permettono l'attivazione dei servizi di prestito e consultazione.
Solo nelle basi dati di polo esistono quindi le informazioni relative agli esemplari e, di conseguenza, anche quelle pertinenti ai possessori/provenienza eventualmente collegati. I dati relativi sono disponibili e visibili unicamente in applicazioni di poli e biblioteche che abbiano curato autonomamente sia la raccolta dei dati che la realizzazione di un'apposita interfaccia per la ricerca. In Italia se ne annoverano diverse, a volte nate per esigenze di servizio a disposizione del personale interno, ma che hanno subito suscitato l'interesse presso un pubblico più vasto e quindi aperte a tutti.
Già da tempo nell'Opac di polo messo a disposizione dall'ICCU, e anche in alcuni altri Opac commerciali, è possibile, in ricerca avanzata, selezionare il campo Possessore dal menu a tendina. Ricordiamo, ad esempio, il catalogo del Censimento regionale delle edizioni del XVI secolo dell'Emilia-Romagna[9] (Fig. 2), quello del catalogo del Polo degli Istituti culturali di Roma IEI[10] (Fig. 3) e del catalogo del sistema bibliotecario padovano[11] (Fig. 4).

[9] Polo del Censimento regionale delle Edizioni del XVI secolo dell'Emilia-Romagna (CER): <https://polocer.regione.emilia-romagna.it/opac/.do>.
[10] https://cloud.sbn.it/opac/IEI/02/ricercaSemplice.
[11] Polo SBN Universitario veneto (PUV): <https://galileodiscovery.unipd.it/discovery/search?vid=39UPD_INST:VU1>.

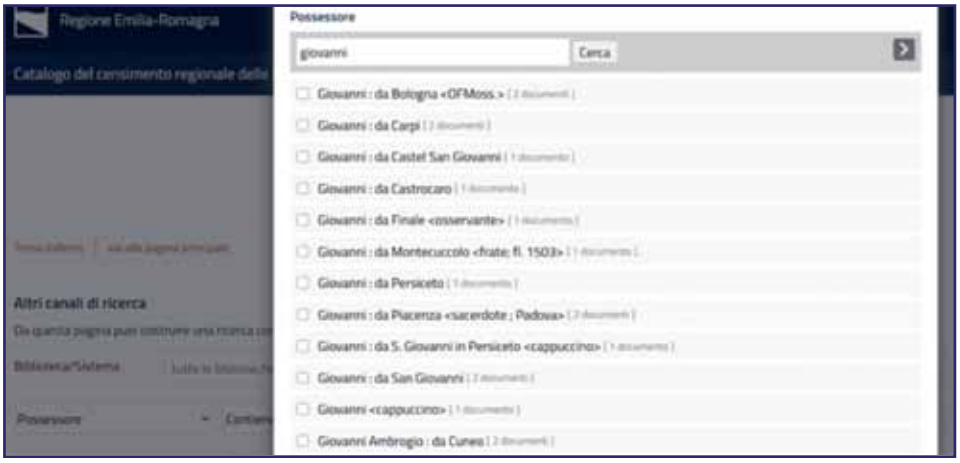

Figura 2. *La lista dei possessori a fronte di una ricerca nel canale Possessori nell'Opac del Polo CER*

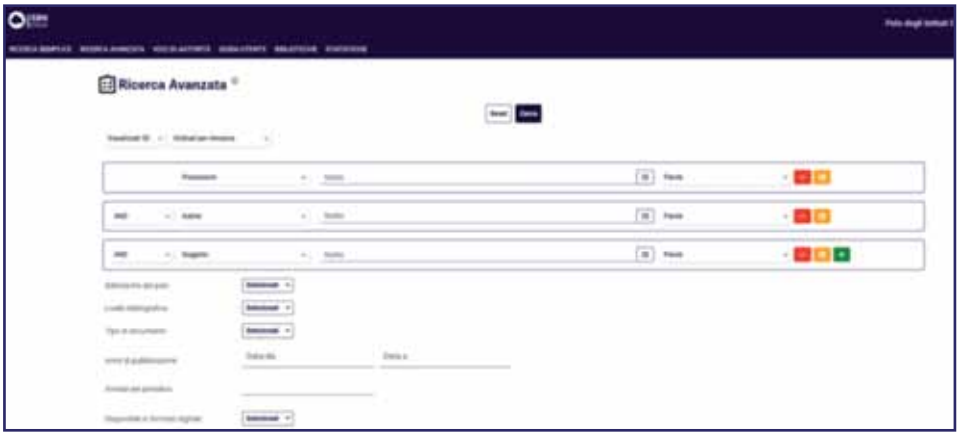

Figura 3. *La ricerca dei possessori nell'OPAC del Polo degli Istituti culturali di Roma IEI*

Figura 4. *La ricerca dei possessori nel Discovery del Sistema bibliotecario padovano*

La ricerca, tuttavia, si limita al possessore; non è possibile, ad esempio, effettuare ricerche sui fondi di una biblioteca e vederne i possessori collegati.
Accanto a questa tipologia di ricerca all'interno degli Opac locali, vanno segnalati dei siti tematici sull'argomento. Per citare qualche caso significativo tra le molte iniziative presenti in SBN, ricordiamo l'Archivio possessori della Biblioteca nazionale Marciana di Venezia[12] (Fig. 5) e quello della Biblioteca nazionale di Napoli[13] (Fig. 6).

Figura 5. *La maschera di ricerca dei possessori messa a disposizione dalla Biblioteca Marciana di Venezia*

Figura 6. *La maschera di ricerca dei possessori messa a disposizione dalla Biblioteca Nazionale Vittorio Emanuele III di Napoli*

[12] <https://marciana.venezia.sbn.it/la-biblioteca/cataloghi/archivio-possessori>. Orsola Braides — Elisabetta Sciarra, *L'Archivio dei possessori della Biblioteca nazionale Marciana: un database di provenienze*, <https://marciana.venezia.sbn.it/sites/default/files/repositoryfile/pagine/2017/ allegati/ircdl2016paper13_0.pdf>.

[13] <http://www.bnnonline.it/index.php?it/330/archivio-possessori>. Si veda: Simona Pignalosa, *I possessori nella Biblioteca nazionale di Napoli: un Archivio di immagini*, «DigItalia. Rivista del digitale nei beni culturali», 10 (2015), n. 1/2, p. 85-95, <https://digitalia.cultura.gov.it/article/view/1477>.

Come si può immaginare, le possibilità di ricerca sono diverse come diversa è anche la presentazione dei risultati ottenuti. Questo aspetto non agevola la ricerca dello specialista che si vede costretto a consultare molteplici basi dati con diverse modalità, ottenendo risultati spesso non omogenei tra loro.
Date queste premesse, risulta ancora più evidente l'esigenza di una struttura di livello nazionale capace di accogliere i contributi delle diverse realtà SBN in un unico punto di aggregazione e di mettere a disposizione la consultazione dei dati relativi a possessori e provenienze a livello nazionale. L'Indice SBN appare la struttura più idonea a rivestire tale ruolo anche per questa tipologia di dati.
Un altro aspetto gioca a favore dell'utilizzo dell'Indice SBN: la catalogazione partecipata, la regola principale su cui si basa il funzionamento di SBN stesso. Tale metodo risulta al contempo efficiente ed economico per il lavoro dei bibliotecari che partecipano alla rete. In pratica, ogni elemento del reticolo bibliografico (titolo, nome, marca ecc.) viene creato dal bibliotecario che per primo ne ha bisogno per catalogare una determinata risorsa; i catalogatori che successivamente avranno la necessità di utilizzare questo elemento dovranno semplicemente catturarlo e, se necessario, modificarlo e completarlo mettendo questa nuova versione a disposizione di tutta la comunità SBN. Attraverso complessi meccanismi di allineamento, tutti gli elementi che compongono la notizia bibliografica vengono costantemente replicati tra i poli e l'Indice SBN.
Il criterio di condivisione partecipata degli elementi del reticolo riguarda ovviamente solo i dati comuni. Tra questi attualmente non sono presenti gli archivi dei possessori che ciascun Polo, mancando in Indice una struttura che li ospiti, mantiene nella propria base dati locale.
Fanno eccezione, rispetto alla situazione appena descritta, i poli che sono entrati a far parte della piattaforma SBNCloud[14]. In questo caso, in fase di migrazione dal vecchio al nuovo applicativo, l'authority Possessori viene migrato e armonizzato all'interno dell'archivio di authority Nomi/Enti, rendendo possibile una efficace pulizia e normalizzazione tra elementi uguali.

Il progetto

L'idea di sviluppare per Fondi-Possessori-Esemplari una funzionalità apposita non è recentissima: studiosi, ricercatori ma soprattutto bibliotecari in varie circostanze hanno chiesto all'ICCU la disponibilità a livello centrale di questa tipologia di informazioni, il cui recupero era possibile solo attraverso la consultazione delle singole basi dati locali.
Su questa spinta, circa otto anni fa, l'ICCU ha costituito un gruppo di lavoro che ha avuto come scopo quello di verificare la realizzabilità di una linea fondi - possessori – esemplari

[14] SBNCloud è il gestionale open-source sviluppato dall'ICCU, che consente alle biblioteche di avvalersi della tecnologia Cloud-computing e di condividere il proprio patrimonio con la rete SBN. Il software si compone di microservizi che possono essere adottati in base alle proprie esigenze. Importanti evoluzioni riguardano la catalogazione semantica, le acquisizioni, la configurazione dei servizi e la gestione delle sale di lettura. SBNCloud è integrato con un servizio di Teca Digitale (Teca Centrale), che genera automaticamente i metadati associati alle risorse digitali caricate, a partire dai dati di catalogo. <https://www.iccu.sbn.it/it/SBN/sbncloud/>.

a livello centralizzato, effettuando un'analisi di fattibilità e un censimento delle basi dati dedicate, incrementate da biblioteche SBN[15].
Nella fattibilità si è tenuto conto dell'aspetto relativo alla presenza di dati già esistenti a livello locale: elemento rilevante, tendendo conto che la nuova struttura costruita per ospitare i dati nasce "vuota", e per il suo popolamento, che avverrà grazie alla cooperazione dei Poli, saranno necessari dei tempi lunghi. Per questo l'ICCU analizzerà la possibilità di creare una procedura di importazione nell'Indice SBN delle basi dati esistenti, almeno quelle più rilevanti per consistenza e importanza.
Per meglio comprendere come è strutturata la funzione e quali saranno i dati messi a disposizione per la ricerca in Opac, descriviamo in sintesi i tre elementi che, in una sorta di triangolazione, si legano tra loro permettendo di ottenere risultati interconnessi.
Partiamo dall'entità Fondo. La soluzione adottata dall'ICCU, attualmente in fase avanzata di realizzazione, prevede il coinvolgimento diretto dell'Anagrafe delle biblioteche italiane.

Gestione dei fondi nell'Anagrafe delle biblioteche italiane

La base dati dell'Anagrafe delle biblioteche italiane - gestita dall'ICCU, frutto di un importante progetto avviato alla fine degli anni '80 con la finalità di censire tutte le biblioteche esistenti sul territorio italiano - costituisce oggi un importante strumento di conoscenza a livello nazionale dei fondi speciali conservati dalle biblioteche.
La scheda di censimento per la redazione del *Catalogo delle Biblioteche d'Italia*[16] - che costituì la prima forma di pubblicazione dei dati raccolti - prevedeva pochi elementi identificativi del Fondo, indicandone sinteticamente la denominazione e la descrizione (breve indicazione del suo contenuto). La ricercabilità dei fondi era garantita attraverso un indice per "denominazione".
Con la reingegnerizzazione della base dati avvenuta nel 2000, l'Anagrafe delle biblioteche italiane è divenuta consultabile in rete[17] e, tra gli altri interventi, è stato esteso anche il profilo descrittivo dei Fondi speciali, per cui sono stati aggiunti alla denominazione e descrizione anche la classificazione Dewey, le condizioni di disponibilità del fondo da parte dell'istituto conservatore, l'indicazione del catalogo/inventario ed eventuale URL della risorsa.
I Fondi speciali registrati in Anagrafe, descritti da questi elementi – la maggior parte dei quali obbligatori per la registrazione sulla base dati – ad oggi, sono quasi 10.000. Si tratta di fondi librari, ma anche archivistici, di diversa natura e tipologia, conservati da biblioteche di pubblica lettura, di conservazione, specializzate e di ricerca, biblioteche pubbliche o biblioteche private. I fondi, pervenuti nelle istituzioni bibliotecarie a seguito di lasciti, donazioni, depositi o acquisti, sono costituiti da una pluralità di tipologie di

[15] Flavia Bruni ha parlato del progetto dell'ICCU nell'articolo già citato alla nota 2.
[16] Il primo volume che venne pubblicato fu quello relativo all'Umbria: Ministero per i beni culturali e ambientali - Ufficio centrale per i beni librari e gli istituti culturali, Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche - Regione Umbria, *Catalogo delle biblioteche d'Italia. Umbria*, Roma: ICCU; Milano: Editrice Bibliografica, 1993.
[17] https://anagrafe.iccu.sbn.it.

materiali, come libri, documenti, fotografie, ma anche oggetti di natura museale prodotti, ricevuti e raccolti da istituzioni, persone o famiglie.
Non è semplice una classificazione delle diverse tipologie censite ma, considerando il nucleo più rappresentativo[18], possiamo distinguere *collezioni e raccolte* di natura bibliografica legate a un possessore, che individuano un insieme organico di esemplari collezionati da una persona, una famiglia o una istituzione, oppure risultato dell'attività dello stesso istituto conservatore che ha raccolto e organizzato le copie con un criterio specifico, come ad esempio il valore storico-artistico, la tipologia documentaria o un tema o soggetto particolare; *fondi personali*[19], vale a dire biblioteche e archivi d'autore e di persona, raccolte ibride che testimoniano l'attività intellettuale di personalità significative del mondo della cultura, delle professioni e delle arti. In modo marginale, sono presenti anche *fondi archivistici* legati all'attività di un soggetto produttore.
Esula da questa classificazione la generica indicazione di "fondo antico"[20], valore utile a indicare la quantità di patrimonio antico conservato dalla biblioteca.
Con la finalità di valorizzare questo vastissimo e altresì eterogeneo complesso di Fondi speciali descritti sulla base dati, si è dato avvio nel corso del 2024 a un progetto specifico di arricchimento ed estensione del profilo descrittivo della scheda dell'Anagrafe delle biblioteche italiane. Lo scopo è stato quello di potenziare le informazioni già a disposizione, fornendo una scheda descrittiva del Fondo che potesse assumere la funzione di completare i dati presenti nell'Indice SBN, facendo emergere le relazioni tra il complesso e il possessore/collettore, ma anche le connessioni tra i documenti che ne fanno parte.
L'importanza a livello catalografico di un modello di descrizione di intere collezioni, che possa accogliere informazioni generali e d'insieme sul complesso documentario – informazioni spesso demandate alle pagine web delle biblioteche o desumibili parzialmente nei cataloghi dalla descrizione delle singole copie – è stata spesso sottolineata[21] nell'am-

[18] I fondi registrati sono prevalentemente fondi di natura bibliografica. Si tratta di circa 7.000 fondi librari/collezioni, 800 fondi archivistici, 2.000 fondi misti. Questa prima classificazione – effettuata sulla base di un'attribuzione automatica basata sulle informazioni già disponibili sui fondi - sarà verificata e oggetto di revisione. I fondi "misti" - che non è stato possibile ricondurre alle altre tipologie - sono stati individuati per esclusione.

[19] I Fondi personali sono «complessi organici di materiali editi e/o inediti raccolti e/o prodotti da persone significative del mondo della cultura, delle professioni e delle arti prevalentemente dalla seconda metà del XIX secolo in poi». Definizione tratta da: Associazione italiana biblioteche. Commissione nazionale biblioteche speciali, archivi e biblioteche di autore, *Linee guida sul trattamento dei fondi personali*, versione 15.1. 31 marzo 2019, <https://www.aib.it/documenti/linee-guida-sul-trattamento-dei-fondi-personali/ >.

[20] Sulla distinzione terminologica tra "fondi antichi" e "collezioni speciali" (assimilabili a "Fondi speciali"), si rimanda a Lorenzo Baldacchini e Anna Manfron nel saggio *Dal libro raro e di pregio alla valorizzazione delle raccolte*, in: *Biblioteche e biblioteconomia: principi e questioni*, a cura di G. Solimine e P. G. Weston, Roma: Carocci, 2015, p. 315-349: 323-326.

[21] Tra gli interventi sull'argomento, si rimanda al saggio già citato di Lorenzo Bianchini e Laura Manfron che sottolineano l'importanza per lo studio del libro antico e dei fondi d'autore delle relazioni del singolo esemplare con altri manufatti dello stesso complesso documentario, *Dal libro raro e di pregio alla valorizzazione delle raccolte*, cit., p. 335-337. Alberto Petrucciani ha rimarcato la necessità per lo studio delle raccolte conservate dalle biblioteche di una "scheda-fondo", «sistematicamente strutturata e integrata o integrabile con le altre (come del resto si fa de plano nei sistemi informativi archivistici)», Alberto Petrucciani, *Dai censimenti bibliografici alla storia della cultura e della società*, cit., p. 257-270.

bito degli studi sul libro antico per la ricostruzione delle raccolte, ma anche in ambito moderno e contemporaneo per i fondi personali, dove è necessario conservare le relazioni tra la raccolta libraria e materiale di altra tipologia non descritto dal catalogo, come documenti d'archivio, carteggi ma anche oggetti prodotti o appartenuti al possessore.
L'utilità catalografica di una descrizione a livello del fondo è riconosciuta anche nell'ambito di materiali speciali descritti in SBN, in particolare le fotografie. Questi materiali sono spesso descritti complessivamente come fondi fotografici: si tratta di cospicue collezioni e raccolte di fotografie realizzate dagli anni Trenta del XIX secolo ai nostri giorni, pervenute alle biblioteche a seguito di lasciti, donazioni, depositi o acquisti e riconducibili a diverse tipologie[22] che, sulla base delle modalità di aggregazione e sedimentazione, possono essere assimilabili a raccolte di natura bibliografica o a fondi archivistici.
In mancanza di un modello consolidato di scheda descrittiva del fondo, almeno per quanto riguarda l'ambito bibliografico, si è provveduto a integrare la struttura dati dell'Anagrafe delle biblioteche italiane sulla base delle indicazioni che sono state elaborate in seno a importanti progetti nazionali[23] e in linea con lo standard archivistico ISAD (G)[24], per quanto attiene le componenti descrittive essenziali del fondo, così da favorire l'interoperabilità dei dati e la possibilità di scambio con altri sistemi, anche afferenti ad altri domini, in considerazione dello sviluppo di piattaforme digitali integrate che possono mettere in relazione informazioni provenienti da diversi cataloghi[25] o fonti differenti del patrimonio culturale.
La scheda, che parte dalla struttura già esistente del Fondo speciale dell'Anagrafe delle biblioteche italiane, è stata articolata in modo da consentire la descrizione di fondi librari ma anche archivistici, che spesso sono conservati e gestiti dalle Biblioteche, così da fornire agli istituti uno strumento unico per la loro descrizione. Nella scheda Fondo di Anagrafe si potrà dare indicazione anche di materiale non gestito da SBN, come documenti archivistici o manoscritti, spesso collegati allo stesso fondo.
Con lo sviluppo sono stati inseriti numerosi attributi finora non disponibili, in aggiunta a quelli già esistenti (denominazione, classificazione Dewey, disponibilità, presenza catalogo/inventario):

[22] Tra queste si trovano: collezioni private, archivi di persona (di letterati, storici dell'arte e artisti in genere), archivi aziendali, archivi di studi e ditte di fotografia, archivi di fotogiornalismo, fototeche d'arte, archivi istituzionali, archivi di storia locale. Per la loro descrizione si rimanda alle *Linee di indirizzo per i progetti di digitalizzazione del materiale fotografico*, pubblicate dal gruppo di lavoro Gruppo di lavoro ICCU sulla digitalizzazione del materiale fotografico, Gennaio 2004,
<https://www.iccu.sbn.it/export/sites/iccu/documenti/Linee_guida_fotografie.pdf>.

[23] Tra questi documenti si segnalano le già citate *Linee guida sul trattamento dei fondi personali*: <https://www.aib.it/documenti/linee-guida-sul-trattamento-dei-fondi-personali>. Per la "scheda fondo" si segnala anche il modello di *Scheda di rilevazione dei fondi librari* proposto dalla Regione Toscana, pubblicata da Paola Ricciardi — Maria Cecilia Calabri, *Le biblioteche d'autore nel censimento dei fondi librari della Regione Toscana: tipologie e localizzazioni*, in: *Collezioni speciali del Novecento: le biblioteche d'autore: atti della Giornata di studio (Firenze, Palazzo Strozzi, 21 maggio 2008)*, Firenze: Polistampa, 2009, p.75-106.

[24] https://icar.cultura.gov.it/standard/standard-internazionali/isad-g.

[25] A proposito di un catalogo "integrato" si rimanda al saggio di Alberto Petrucciani, *Dai censimenti bibliografici alla storia della cultura e della società,* cit.

– l'identificazione con codice univoco (collegato all'ISIL[26] della biblioteca);
– la descrizione/storia - campo ampliato fino a 5000 caratteri;
– una classificazione per tipologia per distinguere i fondi librari (collezioni speciali, raccolte ecc.) dai fondi archivistici e dai fondi "misti" (contenenti oggetti bibliografici, documenti d'archivio e altre tipologie anche non riconducibili a SBN);
– la relazione con un soggetto produttore o un possessore ente-persona-famiglia (registrato come "provenienza"), descritto secondo le norme dell'Authority Nomi/Enti SBN[27] a cui è possibile associare, se esistente, il permalink della scheda SBN e delle note storiche;
– la relazione con altri fondi conservati dalla stessa o da altre biblioteche;
– la copertura cronologica del fondo;
– la consistenza;
– l'indicazione delle prevalenti tipologie documentarie (come manoscritti, periodici, risorse cartografiche, risorse grafico-iconiche, risorse musicali, risorse elettroniche ecc.);
– la modalità di acquisizione;
– la bibliografia di riferimento.

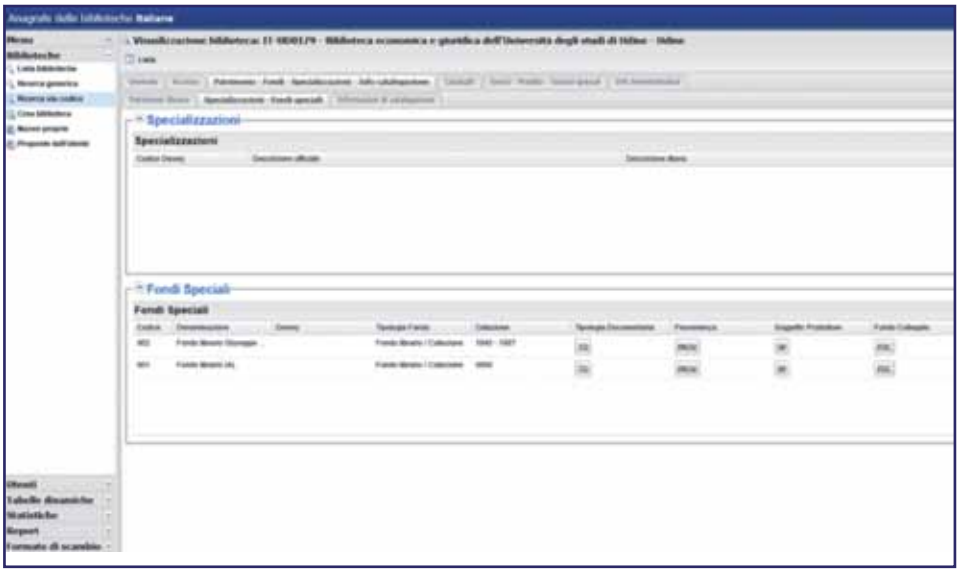

Figura 7. *L'elenco dei Fondi speciali nel gestionale dell'Anagrafe delle Biblioteche Italiane*

[26] Il codice ISIL (International Standard Identifier for Libraries and related organizations) è l'identificativo standard internazionale conforme alla norma ISO 15511 per le biblioteche e le organizzazioni collegate come archivi e musei, e viene assegnato dall'ICCU - Agenzia nazionale per l'Italia riconosciuta dall'ISIL Registration Authority - a tutti gli istituti che si registrano in Anagrafe. Il codice è un requisito necessario per l'identificazione univoca della biblioteca, ed è adottato per garantire l'interoperabilità dei sistemi con SBN e ILL-SBN.

[27] https://www.iccu.sbn.it/it/normative-standard/norme-per-la-catalogazione-in-sbn/.

MODUS

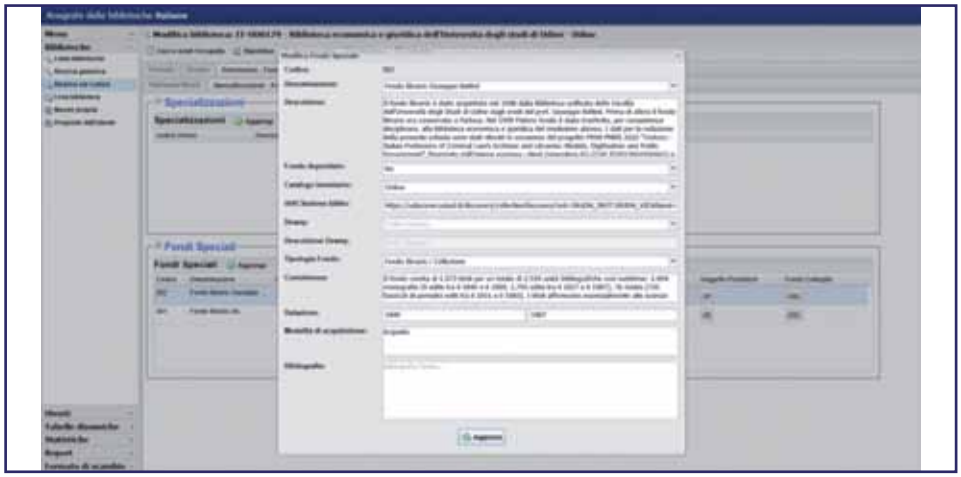

Figura 8. *La scheda Fondi speciali nel gestionale dell'Anagrafe delle Biblioteche Italiane*

L'introduzione di una scheda fondo, strutturata e ricca di informazioni, è un importante passo avanti e una spinta alla valorizzazione dei fondi conservati nelle biblioteche italiane, la cui gestione è spesso limitata e frammentata nei diversi sistemi locali. L'arricchimento della base dati dell'Anagrafe delle biblioteche italiane potrà avvenire grazie al contributo della comunità SBN, ma anche dei singoli istituti censiti che potranno inviare a livello centrale informazioni più dettagliate sui fondi conservati. Lo sviluppo consentirà al ricercatore di interrogare complessivamente i fondi esistenti a livello nazionale sulla base dati, utilizzando numerosi filtri di ricerca, tra quelli disponibili nell'avanzata del sito di Anagrafe (denominazione, tipologia, descrizione, provenienza/soggetto produttore), con la possibilità di restringere l'indagine anche per area geografica o per datazione, senza dover ricorrere a frammentate esplorazioni su diversi sistemi o piattaforme, difformi per la struttura dei dati.

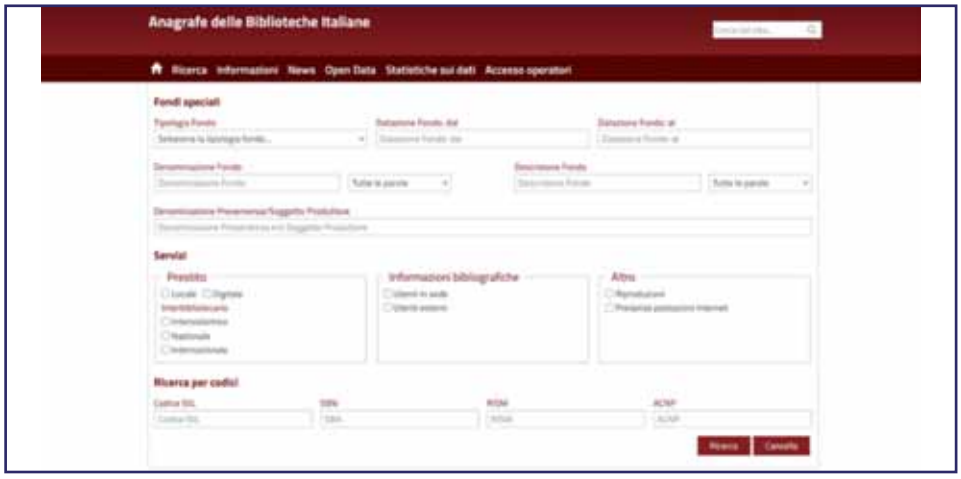

Figura 9. *La ricerca avanzata sui Fondi speciali sul sito dell'Anagrafe delle biblioteche italiane*

La ricerca sui Fondi registrati da Anagrafe è disponibile anche attraverso la il portale Alphabetica[28], mediante il canale "Biblioteche".

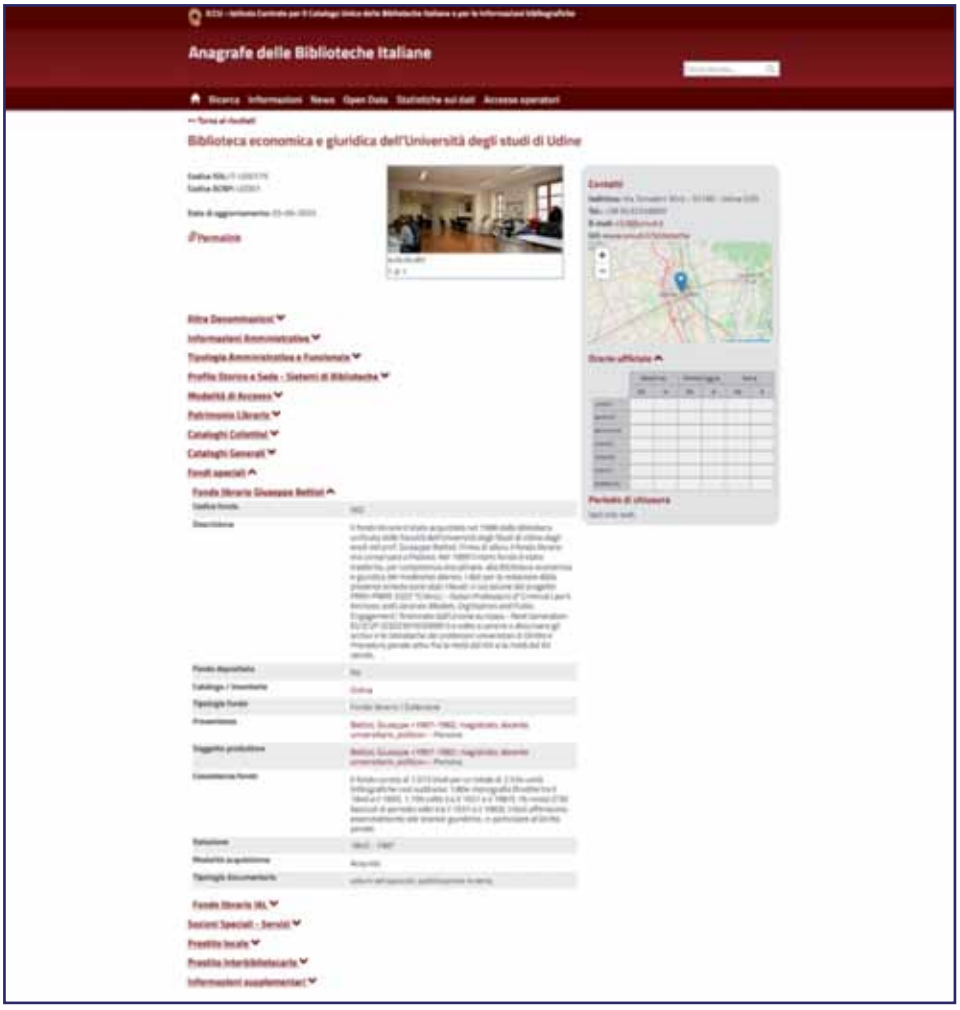

Figura 10. *Il dettaglio di un fondo speciale pubblicato sull'Anagrafe delle Biblioteche Italiane*

Grazie all'intervento evolutivo, ogni Fondo non è più un semplice elemento descrittivo della scheda della biblioteca, ma è identificato in modo univoco da un codice, composto dal codice ISIL seguito da un numero progressivo, aspetto indispensabile per la sua riconoscibilità da parte dell'Indice SBN, ma anche di altri sistemi.
Tutti gli elementi descrittivi relativi ai fondi gestiti da Anagrafe sono rilasciati con licenza CC0 sotto forma di Open Data, aggiornati quotidianamente, in formato XML e disponi-

[28] https://alphabetica.it/web/alphabetica.

bili nella pagina dedicata del sito[29]; sono inoltre pubblicati anche come LOD, insieme agli altri dati relativi alle biblioteche, sul portale www.dati.gov.it. Come per gli altri sistemi gestiti dall'ICCU, è possibile interrogare e recuperare i dati anche attraverso API.

Figura 11. *I fondi sul canale Biblioteche di Alphabetica*

Nel progetto realizzato dall'ICCU le schede Fondo, così arricchite, continueranno a essere gestite, ovvero create e aggiornate, attraverso la base dati dell'Anagrafe delle biblioteche italiane, e verranno messe in relazione ai record bibliografici nell'Indice, con un potenziamento della ricerca nell'Opac SBN, dove l'utente potrà individuare tutti gli esemplari collegati a un fondo speciale.
Con lo sviluppo, i fondi censiti sulla base dati verranno caricati sul data base di Indice 3 con apposita procedura *batch* che inserirà le nuove schede o le aggiornerà a cadenza regolare, mettendo a disposizione di ciascuna biblioteca i dati collegati ai fondi gestiti.

Possessori/Provenienza

Nella nuova funzionalità i nomi utilizzati per designare un possessore o una provenienza fanno parte integrante dell'archivio Nomi/Enti, con una propria scheda di autorità. Si caratterizzano come possessore/provenienza esclusivamente quando posti in relazione con un esemplare o un fondo. Nel rispetto del formato UNIMARC[30] tale legame viene espresso attraverso i codici di relazione 320 *donatore*, (in SBN *provenienza*), e 390 *possessore precedente (*in SBN *possessore)*.
Da notare che lo stesso nome/ente può essere utilizzato indifferentemente sia in relazione ad una manifestazione (quindi alla notizia bibliografica) che all'esemplare, dichiarandone lo specifico livello di autorità e utilizzando il codice di relazione opportuno. Attraverso l'utilizzo di una nota può essere dettagliato il tipo di relazione che lega l'esemplare al nome associato o altre informazioni ritenute importanti.

[29] https://anagrafe.iccu.sbn.it/it/open-data/.
[30] https://www.ifla.org/units/unimarc-rg/.

Nella gran parte dei casi un fondo è associato ad uno o più nomi o ad un ente. Nella funzione viene data la possibilità di esplicitare tale relazione attraverso un legame tra un fondo e uno o più nomi, sempre descritti preventivamente in Authority Nomi/Enti e collegati attraverso il codice di relazione 060, *nomi associati*.

Gli Esemplari

La parte più consistente del progetto riguarda l'inserimento a livello centrale dei dati relativi all'esemplare. In Indice SBN il livello di rappresentazione dei documenti è quello della notizia bibliografica, non delle copie fisiche possedute dalle biblioteche; queste informazioni, come abbiamo detto, risiedono sulle basi dati locali. Si è dunque resa necessaria l'introduzione in Indice di un livello ulteriore, quello dell'esemplare, integrando i dati di ciascuna specifica copia nelle localizzazioni legate alle notizie bibliografiche. Ogni biblioteca potrebbe, infatti, possedere più di una copia relativa alla stessa notizia bibliografica.

Non tutti gli esemplari verranno riversati in Indice ma solo quelli che possiedono almeno una relazione con un fondo e/o con un possessore/provenienza; inoltre per ogni esemplare verranno riportati in Indice solo i dati sufficienti a identificare univocamente ogni singola copia. Non ci sarà quindi un riversamento generalizzato di tutti gli esemplari e dei relativi dati presenti nelle basi dati locali.

Per identificare univocamente la copia abbiamo preso in considerazione i dati inventariali e di collocazione che, uniti al codice ISIL della biblioteca, permettono con certezza di individuare l'esemplare. Le copie registrate possono essere messe in relazione al fondo cui appartengono o ad un nome descritto precedentemente nell'Authority Nomi/Enti e legato all'esemplare mediante l'uso degli specifici codici di relazione: 390 *possessore* o 320 *provenienza*.

I tre elementi in gioco

Una volta inseriti tutti gli elementi e le relazioni necessarie alla descrizione di un fondo di biblioteca o le relazioni che intercorrono tra determinati esemplari e i possessori, abbiamo tutte le informazioni per consentire ricerche a livello nazionale sulle possibili interazioni tra fondi, possessori/provenienza ed esemplari.

A esempio, si potranno incrociare i dati tra un fondo e gli esemplari che lo compongono, oppure, a partire da un possessore, ottenere un elenco di esemplari ad esso associati o ancora, partendo da una copia, conoscere i personaggi che lo hanno posseduto.

Conclusioni

La possibilità di interrogare o ricostruire, a partire da un unico punto di accesso, il patrimonio costituente un fondo, a volte smembrato per disparate vicissitudini e conservato in diverse biblioteche, è uno dei principali punti attrattivi di questo progetto che lo rende di grande interesse per studiosi e professionisti.

L'estensione del principio della catalogazione partecipata anche all'archivio possessori e la creazione di ciascuna voce di autorità condivisa a livello centrale favorirà il lavoro dei catalogatori. In questo modo le schede di autorità potranno essere arricchite grazie all'apporto di ciascuna biblioteca anche con dati di difficile reperibilità al di fuori del contesto locale.

Anche per la descrizione dei fondi, la centralizzazione delle informazioni all'interno dell'Anagrafe delle biblioteche italiane potrà favorirne la ricerca.

Gli utenti avranno l'indubbio vantaggio di poter effettuare un'unica ricerca a livello nazionale, anziché ripetere l'interrogazione su cataloghi diversi e dover confrontare a posteriori i risultati ottenuti.

Queste novità rendono maggiormente qualificato il sistema SBN nel suo insieme, in termini di ottimizzazione dei processi e arricchimento dei dati, offrendo un servizio migliore ai propri utenti, siano essi bibliotecari o utenti del catalogo.

Infine il progetto, valorizzando i beni librari, faciliterà la connessione tra entità e dati provenienti da contesti e domini differenti, generando nuove relazioni conoscitive, con una convergenza sempre maggiore tra biblioteche, archivi e musei.

As part of project Indice 3 *relating to the re-engineering and evolution of the National Library Service (SBN) Index, the Central Institute for the Union Catalogue of Italian Libraries (ICCU) has provided for the possibility of centrally managing the collections of names/entities present in SBN libraries, which until now have only been managed at a local level. Specifically, the Institute is developing the* Fondi-Possessori-Esemplari *(Collections-Owners-Copies) project, which links these elements together. The centralisation in the SBN Index of information relating to copies, collections, provenance/owners and mutual links will enable shared management with all SBN libraries through an expansion of the current structure. To date, each SBN hub that processes this type of data does so completely independently, generating a proliferation of dissimilar systems and offering its users a very diverse and partial search option. With this new project, which is important not only for the SBN community, users will have a single point of access for integrated searches on collections, owners and copies.*

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

Early Modern Digital Itineraries: The Italian Core Set

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00144

Rachel Midura
Virginia Polytechnic Institute and State University (Virginia Tech)

Small, cheaply published itinerary books written by professional travelers indicated precisely which routes to utilize, where to stay, which sites to see, and even provided tools for navigating foreign customs, language, and currency in the early modern period. Early Modern Digital Itineraries (EMDigIt) transforms itinerary books printed from the sixteenth through eighteenth centuries into a unique dataset for the study of historical mobility. This piece describes the workflow for semi-automatic transcription, data tidying, and linking historical place names to modern geographic data. It discusses use cases of the Italian core set, which consists of eight itinerary books published in the Italian language and featuring more than four thousand waypoints. This data is freely available for textual, spatial, and network digital exploration and analysis.

If you teach history; read, or write historical fiction; play, or design historical games; you have likely faced a similar set of logistical questions. The process of imagining yourself in the well-worn shoes of an early modern traveler raises questions such as: where could I travel, and how? How much would it cost me? How long would it take? And, trickiest of all, how would I know any of this?

The initial answer is often an unsatisfactory «it depends», or «it's complicated». Travel memoirs and letters are vague on these nuts-and-bolts details, while official legislation and secondary sources range from the overwhelmingly encyclopedic to the narrowly parochial. Writing about historical mobility requires hours tracking sources to answer a single question for an individual, route, or time period. Recent years have brought excellent new work in mobility studies, as scholars broaden the cast of known travelers, or reconstruct the sensory experience of the travel. Yet the editors of one such volume call for increased attention to the «quotidian mechanics», in other words, «how mobility worked (or did not) in a very concrete sense»[1]. Much like reimagining a travel experience, digital tools for spatial analysis and visualization demand greater clarity than a traditional prose account.

This article introduces the Early Modern Digital Itinerary Project (EMDigIt) which brings sources uniquely suited to answering these questions together with the digital capacity

[1] Rosa Salzberg — Paul Nelles, *Movement and Mobility in the Early Modern World: An Introduction*, in: *Connected mobilities in the Early Modern world: the practice and experience of movement*, ed. by P. Nelles, R. Salzberg, Amsterdam: Amsterdam University Press, 2023, p. 8.

MODUS

for multi-scalar analysis and exploration. Small, cheaply published itinerary books written by professional travelers indicated precisely which routes to utilize, where to stay, which sites to see, and even provided tools for navigating foreign customs, language and currency. Itinerary books structured European conceptualization and navigation of space from antiquity through the eighteenth century. They featured tables of routes detailing how to proceed from one city to another as lists of intermediary stops, often glossed with historical sites, postal stations, or border crossings[2].

EMDigIt pilots a semi-automated transcription, data tidying, and visualization workflow, transforming the itineraries into datasets appropriate for textual, spatial, and network analysis. An EMDigIt web platform will facilitate individual inquiries (e.g. «how could I get from Milan to Paris in 1575?») as well as more complex stories about how routes grew, shrank, and shifted over time. This project joins studies of European road networks such as Itiner-e and Viabundus[3], with the additional benefit of selecting for routes that were widely republished, often over long time periods, and consulted by a variety of readers[4]. Merchants, diplomats, pilgrims, tourists, soldiers, and couriers traveled by these arteries of a pan-European route network.

This article accompanies a first major data release, which we call the Italian Core Itineraries. Eight itinerary books published in Italian from 1562-1720 provide 4,500 individual geo-referenced waypoints utilized by early modern travelers. These eight titles include:

– SAB1562: Anonymous, *Le Poste, Necessarie A Corrieri & Viandanti, Per L'Italia Francia, Spagna, & Alemagna Con le Fiere che si fanno per il Mondo,* Brescia: 1562.
– GH1563: Giovanni dell'Herba, Cherubino Stella, *Itinerario delle poste per diverse parti del mondo*, Roma: 1563.
– OC1608: Ottavio Codogno, *Nuouo itinerario delle poste per tutto il mondo*, Milano: 1608.
– FS1610B: Franciscus Schottus, *Itinerario, overo nova descrittione de' viaggi principali d'Italia*, Padova: 1610.
– OC1623: Ottavio Codogno, *Compendio delle poste*, Milano: 1623.
– SASD: Anonymous, *Poste Diverse d'Italia, Alemagna, Spagna, e Francia*, Milano: SD.
– GM1684: Giuseppe Miselli, *Il burattino veridico, overo, Instruzione generale per chi viaggia*, Roma: 1684.
– GV1720: Giovan Maria Vidari, *Il viaggio in pratica*, Napoli: 1720.

[2] The outcomes of the 2023-2025 NEH Digital Humanities Advancement Grant are available in the project white paper: Rachel Midura, *Early Modern Digital Itineraries: Workshops for Data-Driven Approaches to Premodern Travel White Paper*, 2025 <https://apps.neh.gov/publicquery/AwardDetail.aspx?gn=HAA-293210-23>, and <https://github.com/rmidura/EMDigIt/blob/63b5fef0f9085c6c4cbd465dbaeb990a29477396/Early%20Modern%20Digital%20Itineraries%20White%20Paper.pdf>

[3] Bart Holterman et al., *Viabundus Pre-modern Street Map 1.3*, <http://www.viabundus.eu/>.

[4] Rachel Midura, *Itinerating Europe: Early Modern Spatial Networks in Printed Itineraries, 1545–1700,* «Journal of Social History» 54 (2021), n. 4, p. 1023–1063.

MODUS

1. Project Background

The EmDigIt Project began during my time as a senior graduate research fellow at the Stanford Center for Spatial and Textual Analysis (CESTA). I was at work on my dissertation, now adapted into my first book, on the advent of Europe's early modern postal networks[5]. Ottavio Codogno, the postmaster lieutenant of Milan, published two such itineraries: the *Nuovo itinerario* in 1608 (and reproduced in Venice in 1611 and 1616, see Fig. 1), and the *Compendio delle poste* in 1623. These books are well known for their chapters describing the history and the contemporary operations of the Habsburg postal systems run by the Tassis family of postmaster and postmistress generals. However, most of the book consists of semi-structured route tables (Fig. 1) drawn from Codogno's professional knowledge of pan-European transit and commerce. Codogno assured his readers from the front page that these routes were reproduced «not just for secretaries, but for clerics and merchants»[6].

Figure 1. *Page from a Venetian edition of Ottavio Codogno's Nuovo itinerario (1611). Route tables frequently featured observations on borders, scenery, and historical or religious sites of interest. Bayerische Staatsbibliothek (BSB), München, Res/Geo.u. 87*

[5] Rachel Midura, *Postal Intelligence: The Tassis Family and Communications Revolution in Early Modern Europe*, Ithaca (NY): Cornell University Press, 2025.

[6] O. Codogno, *Nuovo itinerario,* cit.

MODUS

Codogno's book is perhaps the best-known example of a much broader genre. I assembled a bibliography of itineraries and a database of route headers that could be found across itinerary books printed in many different languages and countries[7]. This first iteration of the EMDigIt database supported an article, *Itinerating Europe: Early Modern Spatial Networks in Printed Itineraries, 1545–1700*, in which I argue that network methods reveal the only semi-spatial indexing of space. Proximity was one method among many at work in drawing associations and directionalities among routes. These served as mnemonics for space in a largely pre-cartographic world[8].

In summer 2023 I received a Level 1 Digital Humanities Advancement Grant from the United States National Endowment for the Humanities (NEH) to support a series of workshops advancing data-driven approaches to the history of travel. Our goal was to establish a professional community of researchers on premodern, digital, spatial history and explore how geographic information found in primary sources such as letters and journals could be extracted and mapped with EMDigIt to trace the movement of people and goods. Participants explored shared research questions over the course of the year, at the same time as student research assistants helped to extract and refine data from the corpus of early modern itineraries. By doing so, we aimed to bring desired audiences as stakeholders in the early stages of project design[9], presenting the results at a one-day workshop in conjunction with the Digital Humanities Organization conference in Arlington, Virginia in August 2024.

The year-long series of meetings began with only two itinerary books that had been processed to the point of basic geo-reconciliation (*Compendio delle poste* and *Itinerario delle poste*). From this, we built out an adaptable workflow (Fig. 2) for other itineraries, beginning from those with the shared publication language of Italian. Work with student collaborators significantly improved the accuracy and granularity of the data, including extracting new qualitative features and descriptors, such as borders, monasteries, and warnings[10]. Over 2024, we complete the workflow for an additional six Italian-language itinerary books, chosen in consultation with participating scholars for their omnipresence in libraries and private collections across Europe.

[7] Armando Serra, *"Monopolio naturale": di autori postali nella produzione di guide italiane d'Europa, fonti storico-postali tra cinque e ottocento*, «Archivio per la storia postale: comunicazioni e società», (2023), n. 14–15, p. 40–51. Individual books have been reproduced in edited volumes and datasets: Robert Hibberd — Jack B. Owens, *Before Highway Maps: Creating a Digital Research Infrastructure Based on Sixteenth-Century Iberian Places and Roads*, «Bulletin for Spanish and Portuguese Historical Studies», 40 (2015), n. 1; Clemente Fedele — Armando Serra — Marco Gerosa, *Europa Postale*, Bergamo: Museo dei Tasso e della Storia Postale, 2014.

[8] R. Midura, *Itinerating Europe*, cit.

[9] For the project participants see: <https://emdigit.org/>.

[10] With thanks to the Virginia Tech Computational Modeling and Data Analytics Capstone instructors and students.

MODUS

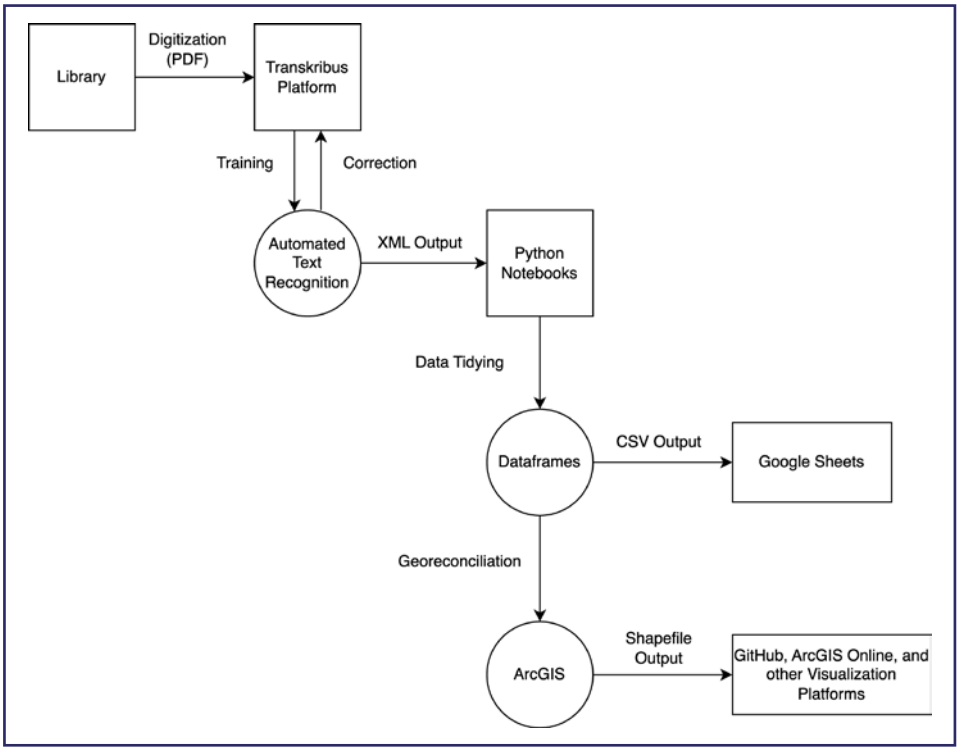

Figure 2. *Generalized model of the EMDigIt Workflow. The Transkribus (READ-COOP)*[11] *platform provides automated text recognition and can be refined through further training on completed itineraries. We then tidy the XML output into data-frames using a set of Python Notebooks. The data can then be exported to Google Sheets as relational tables or imported into ArcGIS for transformation into shapefiles for use in visualization*

We begin from PDFs or archival photographs of the itineraries, uploading them to the Transkribus servers. We then run a neural-network model for text recognition trained on the hand-corrected transcription of *Compendio delle poste* and *Itinerario delle poste*. We use Python to re-segment the data, building and reconciling against a growing project gazetteer. We link out to the GeoNames gazetteer where possible, and calculate approximated coordinates where not based upon journey bearing and distance from prior locations. Our data in turn will be made available through the World Historical Gazetteer (University of Pittsburgh)[12].

Our goal is to release data in stages, with each core set reflecting the most widely republished and utilized itineraries in a given language, as informed by the domain knowledge of project affiliates. With the help of Molly Taylor-Poleskey of Harvard Map Collection, Eva Chodějovská of Masaryk University and a pilot project between Innodata[13] and Harvard Libraries, work is well underway on a set of German-language itineraries.

[11] https://app.transkribus.org/.
[12] Ruth Mostern — Karl Grossner et al, *World Historical Gazetteer*, <https://whgazetteer.org/>.
[13] https://innodata.com/.

The semi-automatic extraction, segmentation, and georeconciliation process will require some refitting to each language, as it relies upon on identifying certain prepositions or key terms. The periodic revision of the workflow, however, also provides staged opportunities to revisit other aspects and to bring in new scholars with pertinent domain knowledge as stakeholders in the project.

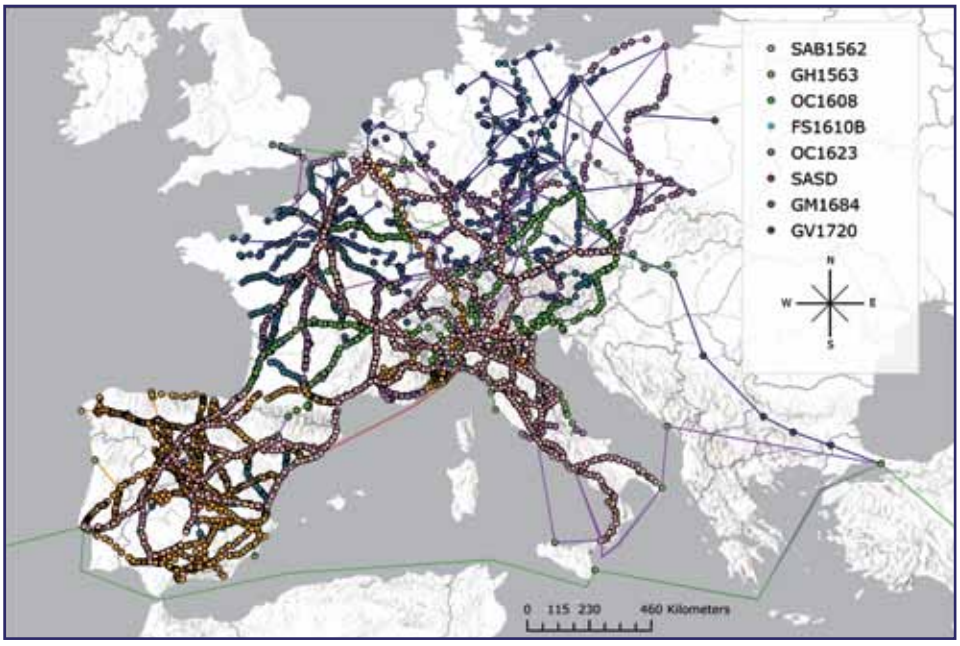

Figure 3. *Map representing all eight itineraries featured in the Italian Core Set of EMDigIt. Each point is an origin, waypoint, or destination that occurs along one or more routes. The itineraries are arranged chronologically, inferring a likely date of publication for the anonymous and undated* Poste Diverse d'Italia, Alemagna, Spagna, e Francia. *Points or routes that occur in multiple itineraries are shown with their earliest title*

The EMDigIt data (Fig. 3) will remain a digital translation of a pre-existing source base of reference literature. While we will continue to publish our workflow and code for scholars at work in other geographic and temporal contexts, we do not intend to open the platform in the style of crowd-sourcing projects. This does mean that EMDigIt will replicate many of the limitations and biases of the sources it transforms. These include (but are not limited to) Eurocentrism, emphasis upon land routes over sea routes, and clear geographic "holes" in the minds of itinerary creators – see, for example, the notable absences in southern Italy, or the Baltic. EMDigIt is not a perfect window on to the experience of travel but rather sheds light on the information that an early modern traveler might have had at hand.
The dynamic nature of our data and intended web platform most closely resembles Orbis: The Stanford Network Model of the Roman World (Stanford University)[14].

[14] Walter Scheidel — Elijah Meeks, *Orbis*, <https://orbis.stanford.edu/>.

This web platform combines cartographic space with experiential factors such as seasonality, elevation, and mode of transport to simulate cost-benefit analysis with contemporary conditions. Users can select points and modes to map their journeys. The site enjoys use by scholars and a wider public for measuring premodern travel, even outside the temporal bounds of the ancient Roman world. The applicability of Orbis for the early modern world is nonetheless limited, as new linguistic, political, and confessional boundaries shaped traveler routes. Orbis has inspired several efforts to produce an «Orbis-in-a-Box», separating methodology and content[15]. We pick up the torch with an important innovation: restricting our data to a delimited and exportable database. Furthermore, EMDigIt will continue to develop in ways that draw from and support projects such as Beyond the Horizon (Moravian Library)[16], Viabundus (Universität Göttingen)[17], and Itiner-e[18].

2. The Italian Core Set

The EmDigIt data consists of four primary objects of analysis: works, editions, routes, edges, and locations. Here I will briefly consider each object in turn as well as their potential for inquiry. Franciscus Schottus' *Itinerario, overo nova descrittione de' viaggi principali d'Italia* serves as a representative example. This work had at least 43 known editions republished across Europe and in many languages. The 1610 edition features 34 routes. These routes include 195 edges, meaning connections among 284 unique locations. While these numbers may seem straightforward, they tell a more complex story.

2.1 Works

First, how do we define a *work,* and an *edition*? There is no need to reinvent the wheel, given the ample scholarship in the history of the book. EMDigIt follow a typical pattern by distinguishing editions by indications that they were printed at a different time, in a different location, or by different publishers. For example, we use FS1610B to refer to the Italian edition published in Venice by Francesco Bolzetta. FS1610A refers to another edition that appeared in Latin in Vicenza the same year[19]. Route tables, however, only appeared in the Italian editions of the work. They had been added by the entrepreneurial Francesco Bolzetta, who continued to publish new editions for several decades. In fact, we see from the digital transformation that publishers regularly intervened in route publication in this way, often copying and

[15] Maxim Romanov — Masoumeh Seydi — James Baillie — Karl Grossner — Rainer Simon — Marfa Vargha, *Orbis-in-a-Box (OIB): Modeling Historical Geographical Networks in ADHO DH Conference at Utrecht University*, 2019, <https://staticweb.hum.uu.nl/dh2019/dh2019.adho.org/programme/book-of-abstracts/index.html>.

[16] Eva Chodějovská et al., *Beyond the Horizon,* <https://beyondthehorizon.mzk.cz/>.

[17] B. Holterman et al. *Viabundus*, cit.

[18] Tom Brughmans — Joseph Guitart — Santiago Muxach — Pau de Soto, *Itiner-e,* <https://itinere.iec.cat/>.

[19] Franciscus Schottus, *Itinerarium Nobiliorum Italiae Regionum, Urbium, Oppodorum, et Locorum*, Vicenza: Petrum Bertellium, 1610.

MODUS

incorporating routes from other publications or making changes where information had presumable become outdated[20].

2.2 Routes

Second, the itineraries are lists of *routes*, with connections between locations either explicitly stated (e.g. «From Rome to Bologna») or simply implied by the sequential logic of a list. We use the Transkribus tool to transform images into XML which preserves important information about the pixel coordinates of text on the original page. As a reference material, itineraries are formulaic by nature: they use prepositions, symbols (such as numbers, asterisks, or printed manicules), and position on the page to signify the purpose of a given text. We can use regular expressions in a Python workflow to distinguish text as a route header, a location name, or a page number.

Symbolic logics do change among titles. Using a dictionary helps to adapt the workflow to a given book but requires periodic human attention. That attention can be productive in its own right; for example, EMDigIt project participants noted that itinerary authors occasionally changed the method by which they measured distance or formatted their text based upon the region in question. This led to new observations regarding where Roman system of measurements predominated over more localized systems, such as the French mile. Furthermore, the digital approach facilitates moving between a qualitative and quantitative consideration. Once location points are geo-referenced, we can measure the modern kilometer distance between them and compare to the original sixteenth or seventeenth century estimation. This in turns provides a dictionary with highly granular data as to what one mile likely meant when used in Provence, Lazio, or Bavaria (Table 1).

Region	Early Modern Italian Mile	Average Distance in Modern Km
Emilia-Romagna	1	1.58
Lazio	1	1.95
Liguria	1	1.48
Lombardy	1	1.84
Piedmont	1	1.82
Veneto	1	1.60
Italy (All Regions)	**1**	**1.84**

Table 1. *While itinerary creators often referred to a unified Italian mile, comparison by modern regions shows the variance from region to region*

[20] It will therefore be valuable to add more editions in future releases. This will support comparison within a given title among editions, and offer greater granularity of change across time to the overall picture of route networks.

MODUS

2.3 Locations

What, then, are the nodes of this network? These are third objects, or the *locations*. The use of a gazetteer, or dictionary of places, is crucial for linking a text object to spatial characteristics such as a set of coordinates and information about region or state. The Geonames Gazetteer (Unxos GmbH)[21] offered a freely accessible starting point, however it is a modern gazetteer that is ill-equipped to help with identifying the location of the «inn of the two sisters», or a location simply designated as «borgo». To this end, we built and published a project gazetteer.

Name variants from itineraries	**Standardized name**	**GeoNames ID**
aborquecque; alberquech; alborqueque; albroqueque; albucheche; alburqueque|berquech; borquech	Alburquerque	2522183
betzanon; bressanon; bressenon; brexanon; brixanon; brixen; brixien; presenon	Bressanone/Brixen	6535887
lengiera; lenguiera; lenguieve	Laneuville au Bois	2793410
mentinsuol; mettinbol; mettinfol; mintecoalt; mirencoalt; mitencoalt; mittewald; montisol	Mezzaselva	3173492
mersulsog; merzhofen; merzuelalag; merzuslalag	Mürzzuschlag	2705
nava de roa; la nava; naue de roa; nava; nava de rona; nava de rova; nava di rova; novederosa; novedrosa	Nava de Roa	3115689

Table 2. *Sample of reconciled place names. The column on the left includes some of the variant names that appear. We can reconcile these with high confidence to a standardized modern place name and authority file (represented by the GeoNames id) based upon their repeated appearances and similar roles across the itineraries*

Reconciliation, deduplication, and reambiguation each play an important role. Our overriding goal is *reconciliation* of each name variant that appears to an authority file, represented here by a GeoNames id and a row in our project gazetteer (Table 2). The *georeconciliation* refers to our ability to fetch spatial data (notably, the latitude and longitude values). To do so, we must often *disambiguate* by drawing connections among name variants. At its simplest, this might mean assuming the Padova that appears in FS1610B is the same as Padova that appears in GH1563. At its most complex, it can mean combining fuzzy matching, experimental mapping, and further research.

[21] https://www.geonames.org/.

We also must often *reambiguate* locations that have been incorrectly collapsed. Examples include the many locations named Villanova, La Venta, or Neumarkt, each a toponym which occurs frustratingly often. This may mean adding a row for which we are not yet able to make an authority file match, but we can still thereby distinguish that a given attestation is *not* the same location as another, matched value.

The eight itineraries feature over 8,800 unique location names. Using a combination of manual review, the fuzzywuzzy package in Python[22], and a custom-built reconciliation tool in Python Bottle[23], we were able to whittle these down to just over 4,500 unique locations. As shown in table 2, many of the name variants strayed far from the modern equivalents. Reading across the itineraries gave important clues for the early modern game of telephone in which Nava de Roa became «Novedrosa», for example. In doing so, we believe we have produced an additional resource for scholars interested in toponymic etymology or linguistics. Variants likely reflect how an Italian author transcribed a name as spoken by locals or other travelers.

In cases where we were not able to match data to a modern gazetteer, the nature of the itinerary still permits well-supported inference using route information about bearing and distance from prior and following identified points. We can provide approximate coordinates using this method, accompanied by a confidence metric. Because many of these locations may no longer exist or go by the same names, the EMDigIt gazetteer also contributes to the long-standing effort to add temporal bounds to geographic data.

The final data object, or *edge*, requires the most inferential approach. We use the term edge here to reflect that each route is a directed network of connected locations (nodes). A route consists of many edges, and edges might be repeated across many different routes. At first glance, the proliferation of waypoints may seem to have solved the teleportation problem of only knowing a start and end point for a given journey. Zooming in, however, reveals that the straight lines between waypoints are equally misleading, zipping imaginary travelers straight across mountaintops or through the middle of lakes.

In network terms, the itineraries give priority to the *nodes* rather than the *edge* connections, which are often left implicit. When the itineraries noted terrain hazards or travel modes, they did so in the same list as other locations. We can use ArcGIS to make more reasonable inferences that consider elevation, terrain, and contemporary borders as edge characteristics (Fig. 4). This is a slow process that often requires the consultation of additional sources, straying farther from the original source material. While this process has only been completed for modern Italy thus far, we hope to offer it as an option for other locations, as it helps to calculate more accurate distance, elevation, and slope attributes for use in determining optimal paths.

[22] https://pypi.org/project/fuzzywuzzy/.

[23] <https://bottlepy.org/docs/dev/>. We took inspiration from the tool designed and published by Ruth and Sebastian Ahnert at <https://github.com/tudor-networks-of-power/code>.

MODUS

Figure 4. *A map represents the route network published in Franciscus Schottus' Itinerario, overo nova descrittione as published by Francesco Bolzetta in 1610. EMDigIt's data construction facilitates moving from the consideration of a full route network to considering single works or editions*

MODUS

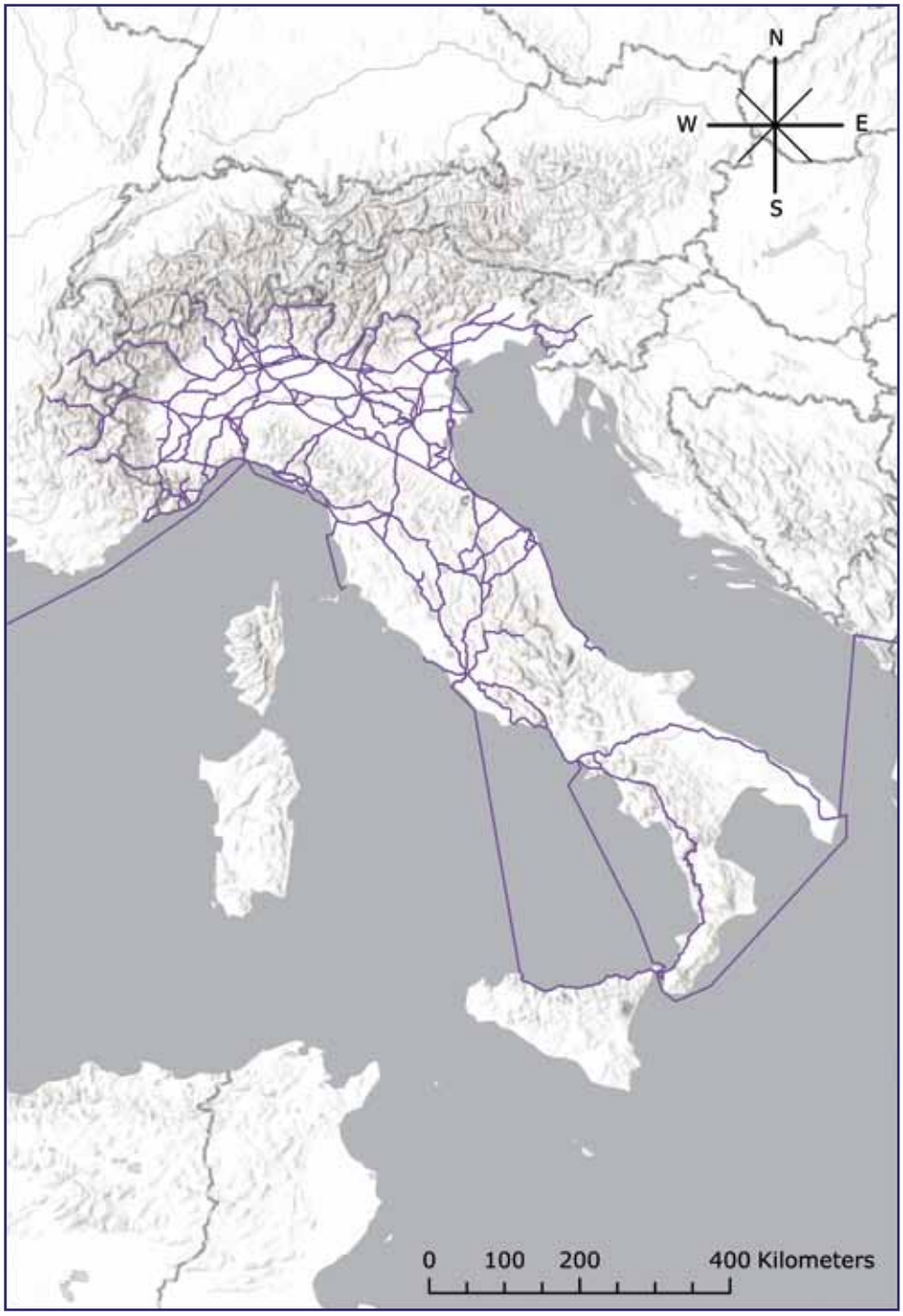

Figure 5. *All eight of the Italian Core Itineraries have been used to create this route network among locations in modern Italy. Instead of straight lines between nodes, the route lines have been further modified to reflect elevation and terrain as well as ancient and modern road networks*

MODUS

3. Insights

In summary, we have introduced several potential use cases for the EMDigIt data:

– The organizing logics of space: as a spatial tradition that predated and cross-pollinated with cartography, the itineraries reflect a "mental map" that is not otherwise easily accessible. The EMDigIt data surfaces organizing the continuities and disruptions of organizing logics; these include the choice to use one system of distance measure or another, designate a place by a given toponym, or reorganize locations into new routes, thereby elevating new locations as origins and destinations.

– The contextualization of individual works or editions: this history of the book approach facilitates the rapid visual comparison of dense reference materials, revealing how authors and publishers have borrowed or modified material over time.

– The enrichment of individual journeys: users seeking answers to logistical questions for a given time, place, or traveler can find contemporary travel guidance to infer where a journey may have gone and experiential details such as its timing, sights and hazards, or modes of travel.

Many of the test cases piloted by project participants involved mutually enriching the EMDigIt data and additional sources, such as *album amicorum* (signature albums), *avvisi* (early newsletters), or postal timetables. EMDigIt remains, at its core, a reference material to provide a comparison and sense of divergence from a norm: as any modern traveler knows, journeys rarely follow their Platonic ideal. EMDigIt benefits from this comparison in turn, as we can refine our calculations of an average pace for different travel modes and determining what factors beyond simple distance went into determining an optimal path. EMDigIt data can help to determine where a traveler might have encountered hazards such as brigands or rough terrain, but it is the traveler accounts that confirm these estimations.

As EMDigIt proceeds through the language core sets, we also look forward to bringing its findings to new audiences. This includes the support of a web platform, but also exploration of its potential use in other creative modes, such as storytelling, games, and augmented reality apps. We take inspiration from projects like HistoryCity (formerly Hidden Cities)[24] which uses spatial data to link travel experiences across the centuries within urban space. We will continue to involve student researchers and collaboration with partners in the U.S. and abroad, as the promise of EMDigIt remains as much in the journey as the destination.

[24] Fabrizio Nevola — David Rosenthal et al., *History City* Apps, <https://historycityapps.org/about/>.

MODUS

Nella prima età moderna, agili ed economiche guide di viaggio realizzate da viaggiatori professionisti indicavano ai viandanti quali percorsi utilizzare, dove alloggiare, quali siti visitare, fornendo anche strumenti per orientarsi tra usanze, lingue e valute straniere. Il progetto Early Modern Digital Itineraries (EMDigIt) *ha lo scopo di trasferire le informazioni contenute nei libri di viaggio stampati dal XVI al XVIII secolo in un dataset, per consentire lo studio della mobilità in questo periodo storico. L'articolo descrive il flusso di lavoro utilizzato per la trascrizione semiautomatica, la pulizia dei dati e il collegamento dei nomi dei luoghi storici ai siti geografici moderni. Sono analizzati come caso d'uso otto libri di itinerari pubblicati in lingua italiana, il set dedicato alle guide italiane, contenenti più di quattromila punti di riferimento. Questi dati sono resi disponibili gratuitamente per l'esplorazione e l'analisi testuale, spaziale e di rete.*

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

Le Digital Humanities alla Biblioteca nazionale universitaria di Torino: sul progetto BINARI e dintorni

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00145

Alessia Grillone
Università degli studi di Torino - Sorbonne Université

*Dal 2022, la Biblioteca nazionale universitaria di Torino (BNUTO) è coinvolta in attività di studio e di valorizzazione incentrate su alcuni manoscritti preziosi, appartenenti ai fondi dell'Istituto, attività condotte in stretta collaborazione con il Centro interdipartimentale di ricerca Digital Scholarship for the Humanities (DISH) dell'Università di Torino. Tra questi lavori vi sono i progetti Il Plinio di Mantegna e BINARI: La Biblioteca Nazionale per Aristotele. Digitalizzazione ed edizione digitale delle traduzioni latine umanistico-rinascimentali dell'Etica Nicomachea di Aristotele nei codici della Biblioteca Nazionale di Torino, entrambi finanziati con fondi PNRR. In particolare, questi si propongono di offrire al pubblico l'edizione digitale di codici che trasmettono le versioni latine dell'*Etica Nicomachea *(EN) aristotelica eseguite da Leonardo Bruni (ms. D V 33, E IV 36, G III 16, G IV 32 e H V 16) e da Giovanni Argiropulo (ms. E III 25), più il celebre codice miniato della* Naturalis Historia *di Plinio il Vecchio (ms. J.I.22-23), testimonianza della ricezione dell'arte di Andrea Mantegna. Il presente contributo si propone di illustrare i lavori su Aristotele e Plinio, descrivendone le varie fasi allo scopo di sottolineare le potenzialità del digitale non solo per la conservazione e la salvaguardia del patrimonio manoscritto, ma anche per lo studio da parte di un pubblico di esperti, favorendo al contempo un approccio interdisciplinare e la divulgazione presso una più ampia cerchia di fruitori.*

Introduzione

Dal 2022, la Biblioteca nazionale universitaria di Torino (BNUTO), in stretta collaborazione con il Centro interdipartimentale di ricerca Digital Scholarship for the Humanities (DISH) dell'Università di Torino (UNITO), è coinvolta in attività di ricerca e di valorizzazione incentrate su alcuni manoscritti di particolare pregio, appartenenti ai fondi dell'Istituto. Questi lavori sono di varia natura: vi è il progetto Tesori del Piemonte, coordinato dal prof. Roberto Rosselli del Turco (UNITO), che riguarda la BNUTO per l'edizione digitale del codice delle epistole di Cicerone J V 34, realizzata dalla dott.ssa Federica Lazzerini (UNITO), e lo studio del manoscritto L.I.4, contenente la traduzione francese del *Romuleon* di Benvenuto da Imola, oggetto della tesi di laurea di Carolina Crespi (UNITO) con il sostegno e il patrocinio della Fondazione Cecilia Gilardi. A queste iniziative si affiancano due progetti, Il Plinio di Mantegna e BINARI: La Biblioteca Nazionale per Aristotele. Digitalizzazione ed edizione digitale delle traduzioni latine umanistico-rinascimentali dell'*Etica Nicomachea* di Aristotele nei codici della Biblioteca

MODUS

Nazionale di Torino, entrambi finanziati con fondi PNRR e condotti nell'ambito del mio dottorato di ricerca presso l'Università di Torino, sotto la supervisione del prof. Ermanno Malaspina (UNITO) e del dott. Fabio Uliana (BNUTO). In particolare, questi due progetti si propongono di offrire al pubblico le pagine più significative del celebre codice miniato della *Naturalis Historia* di Plinio il Vecchio (ms. J.I.22-23), nonché l'edizione digitale di codici che trasmettono le versioni latine dell'*Etica Nicomachea* aristotelica eseguite da Leonardo Bruni (ms. D V 33, E IV 36, G III 16, G IV 32 e H V 16) e da Giovanni Argiropulo (ms. E III 25).

Il presente contributo si propone di illustrare nel dettaglio i miei lavori su Plinio e Aristotele, allo scopo non solo di sottolineare le potenzialità del digitale nell'ambito della conservazione dei beni culturali e degli studi umanistici in generale, ma anche di riflettere sulle implicazioni pratiche del digitale, quali l'acquisizione di determinate competenze in vari campi e la creazione di reti tra enti e discipline differenti.

Digitalizzare: perché

Prima di addentrarmi nella descrizione dei progetti, ritengo sia opportuno ribadire le ragioni sottese alla scelta di intraprendere un'operazione di digitalizzazione[1]. Queste rispondono essenzialmente a due obiettivi, l'accessibilità e la fruibilità da un lato e la salvaguardia dall'altro: la digitalizzazione, infatti, consente all'utente di consultare il manufatto senza alcuna limitazione legata a fattori fisici e/o economici contingenti, garantendo e promuovendo così lo sviluppo della cultura, della ricerca scientifica e della tecnica. Il digitale, inoltre, permette una più ampia fruizione senza alcun pericolo per il documento stesso, dal momento che il numero di casi in cui l'oggetto fisico è maneggiato si riduce, e rende possibile la trasmissione del manufatto nello stato attuale alle generazioni future anche in caso di deperimento o scomparsa del bene. La digitalizzazione, dunque, fa sì che salvaguardia e fruizione si integrino e si potenzino a vicenda, a vantaggio sia degli studi sia della tutela e della promozione del patrimonio delle biblioteche[2].

[1] Cfr. Ministero della cultura, *Piano Nazionale di digitalizzazione del patrimonio culturale*, Versione n. 1.0 – in consultazione 2022-2023, p. 7-8, <https://docs.italia.it/italia/icdp/icdp-pnd-docs/it/v1.0-giugno-2022/index.html>. Per altri casi di recenti progetti di digitalizzazione di archivi e biblioteche si vedano ad esempio: Simona De Lucchi, *L'esperienza di digitalizzazione delle collezioni librarie del Sistema Bibliotecario dell'Università di Firenze: Impronte digitali*, «DigItalia. Rivista del digitale nei beni culturali», 17 (2022), n. 2, p. 107–117, <https://doi.org/10.36181/digitalia-00055>; Emilia Di Bernardo — Marilena Maniaci — Nina Sietis — Nicola Tangari, *MeMo — Memory of Montecassino: Un sistema digitale integrato per la conservazione, la fruizione e la valorizzazione del patrimonio manoscritto dell'Abbazia di Montecassino*, «DigItalia. Rivista del digitale nei beni culturali», 18 (2023), n. 1, p. 113–128, <https://doi.org/10.36181/digitalia-00063>.

[2] Sui benefici del digitale si veda il fondamentale contributo di Gianfranco Crupi, *Biblioteca digitale*, in: *Biblioteche e Biblioteconomia: principi e questioni*, a cura di G. Solimine e P. G. Weston, Roma: Carocci, 2015, p. 373-417; su questo tema si vedano anche gli articoli più recenti di Renaud Milazzo, *La prima campagna di digitalizzazione degli archivi del Venerabile Collegio Inglese di Roma*, «DigItalia. Rivista del digitale nei beni culturali», 19 (2024), n. 1, p. 171–180, <https://doi.org/10.36181/digitalia-00100>, e Pier Matteo Barone, *Between Past and Future: Digital Technologies and the Revolution in Cultural Heritage Preservation*, «DigItalia. Rivista del digitale nei beni culturali», 19 (2024), n. 2, p. 81–99, <https://doi.org/10.36181/digitalia-00103>. Sulle opportunità del digitale per la ricerca scientifica si vedano anche: Alberto Petrucciani, *Il catalogo e la biblioteca digitale: un sodalizio indispensabile per la ricerca*, «DigItalia. Rivista del digitale nei beni culturali», 12 (2018), n. 2, p. 54-62, <https://digitalia.cultura.gov.it/article/

MODUS

A fronte di questi vantaggi e opportunità, è tuttavia evidente che si tratta di un'operazione complessa e onerosa, tale comunque da dover fissare quanto meno delle priorità se non delle limitazioni alle opere potenzialmente candidabili; appare quindi tutt'altro che fuori luogo illustrare quale principio abbia guidato la selezione, all'interno del patrimonio della BNUTO, dei codici sopra citati. Per quanto riguarda le traduzioni di Leonardo Bruni e di Giovanni Argiropulo, la scelta è dipesa dalla loro rilevanza sul piano culturale: entrambe le versioni conobbero un notevole successo editoriale e loro esemplari, manoscritti e a stampa, si diffusero in Italia e all'estero, divenendo così capaci di influenzare le modalità di studio e d'interpretazione dell'*Etica Nicomachea* nell'Europa moderna (XV-XVIII secolo); si tratta, quindi, di testi fondamentali per lo sviluppo del pensiero morale occidentale[3].

view/1865>; Costantino Thanos, *L'evoluzione del ruolo delle Biblioteche Digitali Scientifiche (BDS)*, «DigItalia. Rivista del digitale nei beni culturali», 15 (2020), n. 2, p. 171-173, <https://digitalia.cultura.gov.it/article/view/2640>. Una posizione contraria rispetto al tema della digitalizzazione, invece, è sostenuta da Carlo Federici in *La conservazione del patrimonio bibliografico*, nel già citato *Biblioteche e Biblioteconomia: principi e questioni,* alle p. 523-544 (in particolare 525-526). Pur sostenendo la necessità e l'importanza dei progetti di digitalizzazione riflessioni su alcune problematiche ad essi connessi (quali la cattiva qualità nella scannerizzazione; la mancanza o l'incompletezza di metadati adeguati; la selezione del materiale da digitalizzare; la difficoltà nel reperimento di informazioni bibliografiche complete, organiche e ordinate; la scelta di formati proprietari; una limitata interoperabilità; politiche troppo restrittive nella tutela del copyright e il ruolo giocato dai colossi della *net economy*) si possono trovare in Gino Roncaglia, *I progetti internazionali di digitalizzazione bibliotecaria: un panorama in evoluzione*, «DigItalia. Rivista del digitale nei beni culturali», 1 (2006), n. 1, p. 11-30, <https://digitalia.cultura.gov.it/article/view/311>. In particolare, l'autore si concentra sui progetti collegati a *Amazon Book Search*, a *Google Book Search* - già *Google Print* -, alla *Open Content Alliance* e al progetto di *Biblioteca digitale europea* (sui limiti e i vantaggi di Google Books cf. anche Gino Roncaglia, *Google Book Search e le politiche di digitalizzazione libraria*, «DigItalia. Rivista del digitale nei beni culturali», 4 (2009), n. 2, p. 17-35, <https://digitalia.cultura.gov.it/article/view/276>; Alberto Petrucciani, *La bancarella planetaria e la biblioteca digitale: il punto di vista della ricerca e una possibile agenda per l'Italia*, «DigItalia. Rivista del digitale nei beni culturali», 5 (2010), n. 1, p. 9-32, <https://digitalia.cultura.gov.it/article/view/252>. Riguardo l'importanza di avere una legislazione adeguata e una regolazione delle tecniche e delle pratiche di digitalizzazione e di codifica dei testi elettronici si veda ad esempio Leonardo Candela — Donatella Castelli, *Una teoria fondazionale per le Biblioteche Digitali: il DELOS Digital Library Reference Model*, «DigItalia. Rivista del digitale nei beni culturali», 4 (2009), n. 1, p. 44-82, <https://digitalia.cultura.gov.it/article/view/271>; Rosa Maiello, *Politiche e legislazione dell'Unione Europea per la digitalizzazione del patrimonio culturale*, «DigItalia. Rivista del digitale nei beni culturali», 8 (2014), n. 2, p. 9-23, <https://digitalia.cultura.gov.it/article/view/822>; Giliola Barbero — Roberto Marcuccio, *Manoscritti e metadati nelle biblioteche digitali: progetti italiani e internazionali in due recenti convegni*, «DigItalia. Rivista del digitale nei beni culturali», 11 (2017), n. 1, p. 32-50, <https://digitalia.cultura.gov.it/article/view/1628>. Infine, sulla necessità di formare chi è coinvolto nei progetti di digitalizzazione si veda ad esempio: Maria Guercio, *Le discipline del documento e l'innovazione tecnologica nelle iniziative di formazione degli archivisti*, «DigItalia. Rivista del digitale nei beni culturali», 6 (2011), n. 1, p. 9-28, <https://digitalia.cultura.gov.it/article/view/204>; Gilda Nicolai, *Formare professionisti esperti. Il Master in Catalogazione del manoscritto liturgico medievale: descrizione, digitalizzazione e valorizzazione*, «DigItalia. Rivista del digitale nei beni culturali», 20 (2025), n. 1, p. 69-71, <https://doi.org/10.36181/digitalia-00122>.

[3] Sull'Aristotelismo nel Rinascimento si veda almeno: Charles B. Schmitt, *Aristotle and the Renaissance*, Cambridge (Mass.): Harvard university press, 1983; Luca Bianchi, *Interpréter Aristote par Aristote: parcours de l'herméneutique philosophique à la Renaissance*, «Methodos», 2 (2002), <https://doi.org/10.4000/methodos.98>; *The Cambridge companion to Renaissance philosophy*, ed. by J. Hankins, Cambridge: Cambridge university press, 2007; *The reception of Aristotle's Ethics*, ed. by J. Miller, Cambridge: Cambridge University Press, 2012. Sulla ricezione umanistica dell'*Etica Nicomachea* si vedano almeno Eugenio Garin, *Le traduzioni umanistiche di Aristotele nel secolo XV*, «Atti e memorie dell'Accademia fiorentina di scienze morali La Colombaria» 16 (1951), n. 2, p. 55-104; David A. Lines, *Aristotle's Ethics in the Italian Renaissance (ca. 1330-1650): the Universities and the problem of moral education*, Leiden: Brill, 2002. La bibliografia su Leonardo Bruni è sterminata, pertanto mi limito a fornire solo alcuni titoli, cui rimando per ulteriori approfondimenti: oltre alla monografia *Leonardo Bruni Aretino: Humanistisch-*

MODUS

Il codice pliniano, invece, rappresenta un "fiore all'occhiello" del patrimonio librario della BNUTO: è un manufatto di straordinario valore artistico (Fig. 1), in quanto costituisce una preziosa testimonianza della ricezione dell'arte di Andrea Mantegna. Inoltre, il codice è pregevole anche per il suo contenuto, dal momento che trasmette il testo dell'edizione della *Naturalis Historia* curata dal celebre umanista Guarino Veronese[4]. Del resto, anche tra i manoscritti aristotelici ne figurano alcuni di particolare pregio in virtù del loro apparato decorativo: si tratta dei codici E III 25 e E IV 36, provenienti dalla raccolta libraria di Domenico Della Rovere (1442-1501)[5], nonché

philosophische Schriften mit einer Chronologie seiner Werke und Briefe, herausgegeben von H. Baron, Leipzig-Berlin: Teubner, 1928, per la biografia cfr. la voce: Cesare Vasoli, *Leonardo Bruni,* in: *Dizionario Biografico degli italiani*, vol. 14, Roma: Istituto dell'Enciclopedia italiana, 1972, p. 3-18, <https://www.treccani.it/enciclopedia/bruni-leonardo-detto-leonardo-aretino_(Dizionario-Biografico)/>; sulla versione dell'*Etica*, la sua tradizione manoscritta e il dibattito che essa suscitò si vedano almeno James Hankins, *Repertorium Brunianum: a critical guide to the writings of Leonardo Bruni. Vol. I: A Handlist of manuscripts*, Roma: [Istituto storico italiano per il Medio Evo], 1997; Tomas González Rolán — Antonio Moreno Hernández — Pilar Saquero Suárez-Somonte, *Humanismo y teoría de la traducción en España e Italia en la primera mitad del siglo XVI: edición y estudio de la "Controversia Alphonsiana" (Alfonso de Cartagena vs. L. Bruni y P. Candido Decembrio)*, Madrid: Ediciones Clasicas, 2000; James Hankins, *Notes on Leonardo Bruni's Translation of The Nicomachean Ethics and its receptions in the Fifteenth Century*, in: *Les traducteurs au travail: leurs manuscrits et leurs méthodes. Actes du colloque international organisé par le Ettore Majorana Centre for Scientific Culture (Erice, 30 septembre – 6 octobre 1999)*, ed. by J. Hamesse, Turnhout: Brepols, 2001, p. 427-447; Montserrat Jiménez San Cristóbal, *El valor de un testimonio autógrafo: la versión latina de Leonardo Bruni de la Ética Nicomaquea en el manuscrito XXV F 10 de la Biblioteca Comunale Santa Maria la Nuova de Monreale*, «Cuadernos de Filología Clásica. Estudios Latinos», 32 (2012), n. 1, p. 121-144, <https://doi.org/10.5209/rev_CFCL.2012.v32.n1.39651>; Montserrat Jiménez San Cristóbal, *Algunas notas críticas para la edición de la versión latina de la Etica aristotélica de Leonardo Bruni y la versión castellana de Carlos de Aragón*, «Estudios Clásicos» 152 (2017), p. 153-174. Su Giovanni Argiropulo si veda almeno Giuseppe Cammelli, *I dotti bizantini e le origini dell'Umanesimo. Vol. 2: Giovanni Argiropulo*, Firenze: Le Monnier, 1941, e la voce: Emilio Bigi, *Giovanni Argiropulo,* in: *Dizionario Biografico degli italiani*, vol. 4, Roma: Istituto dell'Enciclopedia italiana, 1962, <https://www.treccani.it/enciclopedia/giovanni-argiropulo_%28Dizionario-Biografico%29/>; in particolare, sul suo insegnamento e la sua relazione con il testo dell'Etica, oltre alla già citata monografia di Lines, cfr. anche Arthur Field, *The Origins of the Platonic Academy of Florence*, Princeton: Princeton University Press, 1988, e Deno J. Geanakoplos, *Constantinople and the West: essays on the Late Byzantine (Paleologan) and Italian Renaissance and the Byzantine and Roman Churches*, Madison: The University of Wisconsin Press, 1989; Daniela Gionta, *Dallo scrittoio di Argiropulo: un nuovo paragrafo della fortuna dell'Etica Nicomachea tra Quattrocento e Cinquecento*, «Studi Umanistici» 3 (1992), p. 5-57. Mancano ad oggi studi sulla sua attività versoria.

[4] Cfr. Michael Reeve, *The editing of Pliny's Natural History*, «Revue d'histoire des textes», 2 (2007), p. 107-179. Per un profilo biografico di Guarino si veda: Alessia Grillone, *La traduzione latina del Nicolces isocrateo di Guarino Veronese: introduzione, edizione critica e commento*, Berlin: De Gruyter, 2023 (in particolare su Plinio p. 21) con relativa bibliografia.

[5] Per la biografia del cardinale Della Rovere si veda la voce: François Charles Uginet, *Domenico Della Rovere*, in: *Dizionario Biografico degli Italiani*, vol. 37, Roma, Istituto dell'Enciclopedia Italiana, 1989, <https://www.treccani.it/enciclopedia/domenico-della-rovere_(Dizionario-Biografico)/>). Sulla sua biblioteca invece si vedano: Gian Carlo Alessio, *Per la biografia e la raccolta libraria di Domenico Della Rovere*, «Italia Medievale e Umanistica», 27 (1984), p. 175-232 (in particolare sul ms. E III 25 cfr. p. 189, 218, mentre sul ms. E IV 36 cfr. p. 222); Ada Quazza — Silvana Pettenati, *La biblioteca del cardinal Domenico Della Rovere: i codici miniati di Torino,* Firenze: Leo Olschki, 1985 (estratto da: *La Miniatura Gotica e il Rinascimento, II: Atti del II Congresso di Storia della Miniatura Italiana, Cortona 24-26 settembre 1982*), in particolare sui ms. E III 25 e E IV 36 cfr. rispettivamente p. 689 e 687); Silvana Pettenati, *La biblioteca del Cardinale Domenico Della Rovere*, in: *Il teatro di tutte le scienze e le arti. Raccogliere libri per coltivare idee in una capitale di età moderna, Torino 1559-1861, Catalogo di mostra: Torino, Archivio di Stato 22 novembre 2011-29 gennaio 2012*, a cura di I. Massabò Ricci et al., Torino: Ministero per i beni e le attività culturali, Direzione regionale per i beni culturali e paesaggistici del Piemonte, 2011, p. 125-128 (in particolare la scheda n. 98 a p. 125 relativa al ms. E III 25).

MODUS

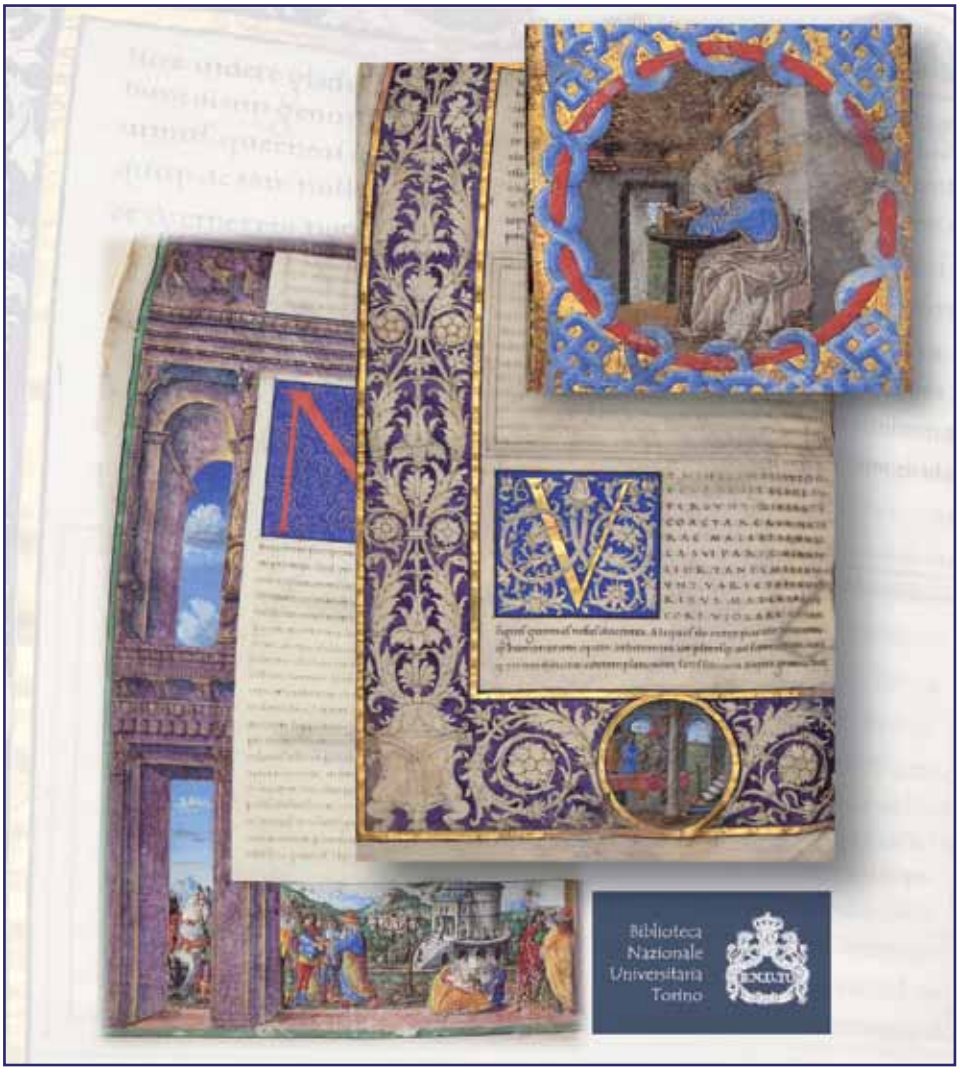

Figura 1. *Particolari del codice miniato della* Naturalis Historia *di Plinio il Vecchio* (Torino, Biblioteca Nazionale Universitaria, ms. J.I.22-23)

dal manoscritto D V 33, miniato dal cosiddetto Maestro delle *Vitae Imperatorum*[6].

[6] Ibidem, p. 61-64, in particolare p. 62-63 (scheda n. 20).

MODUS

Le fasi primarie del processo di digitalizzazione: scansione e metadatazione

Il processo di digitalizzazione di un testo, sia esso a stampa o manoscritto, muove necessariamente dalla trasposizione in forma di immagini delle sue pagine. Per i succitati codici ci siamo avvalsi di un sofisticato scanner piano equipaggiato con una fotocamera dotata di un sensore a pieno formato da 45,4 megapixel. Il ricorso ad uno strumento estremamente specializzato è stato determinante date le peculiarità degli oggetti da convertire in digitale, quali il formato e le dimensioni fuori standard; la delicatezza e la fragilità dei manufatti, caratterizzati da un limitato angolo di apertura e particolarmente sensibili anche alla luce, che impongono seri vincoli sia all'illuminazione da applicare sia al posizionamento sul piano di scansione; la ricchezza dei colori e la morbidezza degli inchiostri, che esigono di operare scansioni ad altissima risoluzione e profondità di colore in tempi accettabili, riducendo al minimo la necessità di interventi correttivi in post-produzione. Ad esempio, nel caso del Plinio, la scansione di un singolo foglio di 41 x 31 cm ha prodotto un file in formato TIFF (compressione senza perdite) tra gli 80 ed i 90 megabyte, corrispondente ad un'immagine di circa 6500 x 4800 pixel, con una risoluzione di 400 punti per pollice e una profondità colore di 24 bit, cui va aggiunto un ulteriore intervento di post-produzione via software per assicurare la piena fedeltà all'originale in termini di colore.

Di contro, un tale oggetto digitale pone dei problemi tutt'altro che trascurabili per quanto concerne sia gli spazi di archiviazione sia, soprattutto, il loro impiego in contesti reali, quale, ad esempio, la fruizione attraverso una app o un sito web. Per ovviare a quest'ultimo inconveniente, a partire dalla scansione originale sono state realizzate in post-produzione copie di lavoro di dimensioni più ridotte tenendo conto del tipo di utilizzo e della modalità di fruizione, senza sacrificare eccessivamente la qualità. A titolo di esempio, le immagini del Plinio caricate sul sito web della BNUTO sono file in formato jpeg del peso di circa 2 megabyte, con una risoluzione di 72 punti per pollice, del tutto adeguate a una visualizzazione via schermo e pienamente compatibili con il tempo di attesa richiesto dal loro caricamento attraverso la rete. Tuttavia, come avrò modo di illustrare più avanti, in un altro contesto si è reso necessario spingersi a livelli di compressione ancora più elevati, nell'ordine di poche centinaia di kilobytes, per poterne garantire l'impiego.

Contrariamente a quanto credono i non addetti ai lavori, però, il processo di digitalizzazione non si arresta, e non si limita, a questo stadio, ovvero alla produzione di immagini del manufatto. Occorre, infatti, prevedere almeno un'altra operazione, ossia la generazione di un insieme di dati indispensabili per l'identificazione, la rintracciabilità e la gestione dell'oggetto digitale, nonché per la descrizione della struttura interna del testo, i cosiddetti metadati[7]. Per i codici qui presentati, i metadati raccolti conservano e restituiscono, tra le altre, informazioni inerenti alla descrizione dell'oggetto fisico: si va dal

[7] Trattasi di un requisito fondamentale in ogni progetto di digitalizzazione: cf. Ministero della cultura, *Piano Nazionale di digitalizzazione del patrimonio culturale*, cit., p. 37.

MODUS

supporto materiale alle dimensioni, dal totale dei *folia* alla valutazione sulla numerazione (se corretta e originale), dalla legatura all'apparato decorativo, al tipo di scrittura e alle sue caratteristiche, alla presenza di note (ad esempio quelle di possesso) e all'identità del copista (qualora conosciuta); inoltre, vi si trovano l'elenco delle opere contenute e notazioni riguardo eventuali restauri subiti. A queste si aggiungono, poi, le indicazioni di carattere storico: si parte da dove, quando e per chi è stato realizzato il manoscritto, fino alla ricostruzione delle vicende e dei passaggi di proprietà che lo hanno condotto alla sua collocazione attuale.

La prosecuzione del processo di digitalizzazione: la trascrizione

Attraverso i progetti Il Plinio di Mantegna e BINARI si è voluto realizzare un prodotto digitale ancora più pratico, versatile e attraente sia per la comunità degli studiosi, i quali potranno disporre di un nuovo strumento per le loro ricerche, sia per il pubblico non specialistico. Per farlo si è reso necessario provvedere ad un'ulteriore operazione: la trascrizione dei testi trasmessi dai vari codici[8].

Anche in questo caso si è trattato di una scelta ponderata, che ha tenuto conto dell'opportunità e dei vantaggi di tale lavoro che può inizialmente apparire come dispendioso, se non addirittura inutile. A determinarla hanno concorso alcune considerazioni: innanzitutto, la trascrizione garantisce una consultazione decisamente più alla portata di chi, da studente[9] o da semplice curioso, entra in contatto con realtà museali e archivistiche. Ma possono beneficiare della trascrizione anche altri studiosi privi di competenze paleografiche o che desiderano acquisirle (per questi ultimi, dunque, la trascrizione si rivela un prezioso ausilio); ad esempio, un manufatto come il Plinio si avvantaggerà dell'opera di trascrizione in quanto lo renderà più fruibile a tutti quegli storici dell'arte che non dispongono di una formazione in paleografia latina.

Una seconda ragione è legata allo stato di conservazione dei manoscritti, una parte non trascurabile dei quali (tra cui anche il Plinio) è rimasta danneggiata dall'incendio scoppiato nel 1904 all'interno della BNUTO[10]; in conseguenza di ciò, alcuni codici risultano parzialmente illeggibili se non mediante il ricorso ad appositi strumenti (come la Lampada di Wood). Per questi esemplari, quindi, la trascrizione diventa indispensabile, in quanto va ad integrare l'immagine nelle sue parti maggiormente o irrimediabilmente compromesse.

---

[8] Su questo tema si veda ad esempio Elisa Bastianello, *Digitalizzazione, trascrizione, citazione: le fonti testuali per le pubblicazioni digitali*, «DigItalia. Rivista del digitale nei beni culturali», 18 (2023), n. 2, p. 162–172, <https://doi.org/10.36181/digitalia-00084>; Stefano Bazzaco, *La trascrizione automatica di documenti a stampa antichi: appunti per un modello di riconoscimento della tipografia in corsivo*, «DigItalia. Rivista del digitale nei beni culturali», 19 (2024), n. 1, p. 63–86, <https://doi.org/10.36181/digitalia-00094>.

[9] Negli ultimi anni, infatti, le scuole sono entrate sempre più in contatto con le realtà museali e archivistiche per la realizzazione dei progetti di PCTO (Percorsi per le competenze trasversali e l'orientamento).

[10] Su questa vicenda e i successivi interventi di recupero del patrimonio si veda il saggio di Franca Porticelli, *Torino 1904. Dopo l'incendio: 1. Gli interventi nell'emergenza e i primi restauri; 2. La decisione di costruire una nuova sede: un'indagine nei documenti d'archivio*, «Scrineum» 17, (2020), n. 1, p. 107-192, <http://www.serena.unina.it/index.php/scrineum/article/view/7964>. Per ulteriori approfondimenti sul tema rimando alla bibliografia di Porticelli.

MODUS

Da ultimo, la trascrizione è funzionale a realizzare l'edizione digitale del codice stesso, consentendo, come vedremo meglio più avanti, di marcare specifici elementi del testo in modo da renderlo un ipertesto.

Riguardo alle modalità di esecuzione, da compito esclusivamente manuale, lento e gravoso, la trascrizione oggi può basarsi sul supporto, ed avvantaggiarsene, di alcuni software innovativi che, in modalità semiautomatica, sono in grado di riconoscere con grande precisione e autonomia testi manoscritti o a stampa a partire da immagini del loro contenuto. Nei progetti che sto curando mi sono avvalsa di Transkribus[11], un software open source sviluppato dal DEA group (Digitalisierung & Elektronische Archivierung) dell'Università di Innsbruck, in collaborazione con altre undici istituzioni accademiche, e gestito da Readcoop, una cooperativa senza fini di lucro amministrata dagli sviluppatori con la partecipazione e il contributo degli utenti che, per utilizzarlo, sono chiamati a farne parte[12]. Grazie al contributo e al sostegno della Fondazione Cecilia Gilardi, ho trascorso, in qualità di *visiting scholar,* due settimane presso l'università austriaca, a diretto contatto sia con il gruppo responsabile del software sia con alcuni ricercatori impegnati nella sperimentazione e convalida di modelli specifici per la scrittura latina di epoca umanistica, con i quali ho collaborato e mantengo la collaborazione.

Transkribus (Fig. 2) rientra nella categoria dei software di riconoscimento automatico del testo nota come HTR (Handwritten Text Recognition), evoluzione di quella OCR (Optical Character Recognition) che ha inaugurato e per lungo tempo costituito l'unica soluzione per i progetti di digitalizzazione. Come l'acronimo suggerisce, la tecnologia OCR si basa sull'addestramento al riconoscimento dei singoli caratteri a partire da immagini del testo da trascrivere ed è ormai disponibile in una folta schiera di software commerciali. Questo approccio si è rivelato funzionale ed efficace per i moderni testi a stampa, mentre ha mostrato evidenti limiti strutturali per i manoscritti, ancor più se danneggiati, dal layout non perfettamente lineare e con tipi di carattere complessi e non uniformi. I software HTR sono stati sviluppati proprio per superare tali limitazioni, facendo largo uso di algoritmi di intelligenza artificiale orientati al riconoscimento non solo di singoli caratteri, ma anche e soprattutto di intere porzioni di testo (linee, aree, parole) previo addestramento. Proprio riguardo a questa fase, Transkribus offre ai suoi utenti l'opportunità di sviluppare in proprio un modello specificamente allenato sulla grafia da trattare, qualora quelli già disponibili risultino poco efficaci. A tal fine l'utente provvede a selezionare e trascrivere manualmente una porzione rappresentativa del testo in modo da tenere conto di ogni possibile variabile (incluse stesura e mano). Come ordine di grandezza occorre prevedere un battente minimo compreso tra le diecimila e le quindicimila parole, ma detto numero è fortemente influenzato dall'ampiezza dello spettro delle varianti da considerare[13], anche se è evidente che l'*optimum* consiste nel mantenere questa base dati la più contenuta possibile. A partire da essa, il software ricava per iterazioni succes-

[11] https://transkribus.org.

[12] Tra i membri figura anche il Centro DISH, presente attraverso il suo attuale presidente, il prof. Ermanno Malaspina.

[13] Sono stati sviluppati modelli che hanno richiesto fino a centomila parole per essere opportunamente addestrati.

MODUS

sive la cosiddetta *ground truth*, cioè le regole attraverso le quali riconoscere il testo da trascrivere (e che applica ad ogni iterazione) e il campione usato per misurare (anche qui ad ogni ripetizione) la risposta del modello rispetto al risultato atteso attraverso il cosiddetto CER (Character Error Rate), ossia l'indice che determina la precisione nonché l'affidabilità del modello stesso. L'addestramento si conclude quando il CER si assesta su un valore stabile costante ovvero quando viene raggiunto il numero massimo di iterazioni[14]. Con un'oculata scelta del testo selezionato per la *ground truth* si arriva ad un valore del CER inferiore all'1% (ovvero una precisione teorica del 99%).

Figura 2. *La trascrizione del manoscritto E IV 36 condotta con il software Transkribus*

Ovviamente l'efficacia del riconoscimento dipende in larga parte dalla qualità delle immagini di partenza e, anche in questo caso, rispetto agli OCR, i software HTR mostrano una maggiore tolleranza riguardo a imperfezioni e disturbi di qualsiasi natura, siano essi presenti già a monte o a valle della scansione. Da ciò deriva che una delle strategie per portare il modello a livello ottimale consiste in un trattamento preventivo delle immagini teso a migliorarne la leggibilità, passaggio che, però, per i codici in oggetto non è stato necessario. Tuttavia, si è

[14] Questo numero non è fisso, ma dipende da come procede la fase di allenamento del modello, in quanto è funzionale ad arrestare il training quando il CER non riesce ad attestarsi su un valore costante. Trankribus consente di fissare due parametri, entrambi facoltativi: il primo, *training cycles*, ha un valore standard 100 che corrisponde al numero massimo di iterazioni, una volta raggiunto il quale l'addestramento si arresta; il secondo, *early stop*, ha un valore standard di 20 e rappresenta il numero minimo di iterazioni da raggiungere, se il CER si è stabilizzato su un valore costante.

scelto di non impiegare Transkribus per i codici G III 16, G IV 32 e H V 16, a causa del loro stato di conservazione, della quantità di testo richiesto dall'addestramento in rapporto a quella complessiva da trascrivere e dalla difficoltà e onerosità insite nella sua selezione in rapporto al risultato ottenibile; per le stesse ragioni non ho adoperato Transkribus per trascrivere le carte miniate di Plinio, ma ho proceduto manualmente, dato il loro numero contenuto (per contro, il lavoro sarebbe risultato troppo oneroso in termini di tempo per i manoscritti sopracitati, che, pertanto, sono stati lasciati, per il momento, privi di trascrizione).
Per tutti gli altri codici (D V 33, E III 25, E IV 36), invece, la trascrizione si è avvalsa di un modello *ad hoc,* da me sviluppato in collaborazione con Stefan Zathammer, membro del Ludwig Boltzmann Institute for Neo Latin Studies, intitolato *Humanist Handwriting 0.4.2*; detto modello ha richiesto circa tre mesi di lavoro ed è stato addestrato con circa 66.000 parole fino a raggiungere un CER pari a 4,1%.
A partire da uno dei modelli disponibili e dalle immagini con il testo da trattare, Transkribus procede con la trascrizione automatica e ne restituisce il risultato sotto forma di file in formato editabile.
Da ultimo, vale la pena segnalare che, a differenza di altri software, Transkribus (insieme con i suoi modelli di riconoscimento e con quelli eventualmente sviluppati dall'utente) risiede e funziona in cloud, perciò non ha bisogno di essere installato sul computer dell'utente e, di conseguenza, ha bassissimi requisiti in termini di risorse hardware (processore e memoria RAM *in primis*).
In base a quanto detto, si evince chiaramente che l'uso di un qualunque software di riconoscimento automatico supporta sì il ricercatore, ma non è in grado di sostituirlo completamente, anche perché sono in capo a quest'ultimo i criteri, e quindi lo stile, cui l'opera di trascrizione si deve uniformare. Nel mio caso, ho scelto di mantenere la *mise en page*, l'ortografia e l'interpunzione dell'originale, ma ho provveduto a sciogliere gli eventuali nessi, le abbreviazioni e i segni convenzionali, conservando solo la *e* con la cediglia (ę). La trascrizione riporta anche le note, marginali e interlineari, come pure tutti gli elementi paratestuali e le correzioni scribali; laddove la lettera capitale era assente, ho provveduto alla sua integrazione.

L'edizione digitale

Come accennato in precedenza, il possesso di un testo editabile ne rende possibile la codifica in forma ipertestuale, aprendo così la strada alla creazione di edizioni digitali di varia tipologia a seconda dello scopo e della natura del testo originale[15]. Per tutte i presupposti sono i seguenti: l'adozione del linguaggio XML (Extensible Markup Language) conforme allo schema di marcatura TEI (Text Encoding Initiative), specificamente indirizzato agli studi letterari, e la disponibilità di un software in grado di restituirne il contenuto informativo.
La codifica nel linguaggio XML-TEI è stata condotta con l'impiego di un editor commerciale (Oxygen v23.1) e ha rappresentato la parte più onerosa del processo di digitalizza-

[15] Cfr. anche E. Bastianello, *Digitalizzazione, trascrizione, citazione*, cit.

MODUS

zione. Di contro, questa impostazione garantisce la realizzazione e la gestione di edizioni digitali diplomatiche, interpretative e critiche, con la possibilità di integrare le prime due con i *markup* tipici della terza e viceversa.

Per visualizzare le varie tipologie (o anche solo i facsimili digitali, cioè le semplici immagini dei manoscritti), la codifica è stata ottimizzata per il software open source EVT (Edition Visualization Technology), sviluppato a partire dal 2013 dal gruppo di ricerca del prof. Rosselli Del Turco (UNITO) tra l'Università di Pisa e il Centro DISH di Torino nell'ambito del progetto Digital Vercelli Book[16] e attualmente disponibile nella versione 2.0[17]. La sinergia con questo gruppo di lavoro ha consentito, da un lato, di sperimentare il risultato della codifica e di beneficiare del supporto per risolvere eventuali problematiche; dall'altro, di suggerire aggiornamenti al software e di indirizzarne lo sviluppo.

Oltre a ciò, EVT si contraddistingue per la sua aderenza allo standard XML-TEI in ossequio al principio ispiratore della riproducibilità e sostenibilità nel tempo delle edizioni digitali costruite in questo modo, oltre che per l'interfaccia semplice e intuitiva dal punto di vista sia dell'editore sia del fruitore.

Per il progetto BINARI (ms. D V 33, E III 25, E IV 36) ho realizzato delle edizioni diplomatiche interpretative[18] (Fig. 3), in cui l'immagine del manoscritto è affiancata dalla trascrizione del contenuto. Il documento XML è stato dotato di un *markup* di tipo descrittivo, ovvero ho annotato le caratteristiche materiali del codice, quali le lettere capitali (in grassetto ed evidenziate in giallo), le rubricature (in rosso), i danni materiali riportati dal manufatto (laddove la lettura è incerta il testo è evidenziato in grigio), le note marginali e interlineari (la cui posizione è riprodotta nella trascrizione) e le correzioni scribali (qualora sul manoscritto siano intervenute più mani, la differenziazione di queste avviene tramite i colori; per gli emendamenti ho anche specificato se la correzione è avvenuta in rasura, tramite cancellazione o ancora per espunzione). Inoltre, la marcatura segnala la struttura del testo, cioè la scansione in libri (l'*Etica Nicomachea* è costituita da dieci libri) e in unità minori (*tractatus*, cioè gruppi di capitoli, e *capitula*, capitoli singoli)[19] e i paratesti anteposti alle due versioni (le epistole di dedica indirizzate a papa Martino V, destinatario della versione bruniana [E IV 36], e a Cosimo de' Medici, cui è offerta la traduzione di Argiropulo [E III 25]; il testo di Bruni è preceduto anche da una *Praemissio* [D V 33, E IV 36], in cui l'Aretino parla delle proprie scelte traduttive in polemica con l'interprete medievale)[20].

[16] Si veda <http://vbd.humnet.unipi.it/>.

[17] Sono in corso i lavori di implementazione per la versione 3.0.

[18] Per il ms. E III 25 in un prossimo futuro sarà disponibile anche un'edizione critica: infatti, nell'ambito della mia tesi di dottorato sto curando il testo critico della versione di Argiropulo.

[19] La divisione in trattati e capitoli vale solo per il testo argiropuleo ed è riconducibile all'autore, mentre la versione bruniana si presenta suddivisa unicamente in libri nel ms. E IV 36 e nel ms. D V 33 in libri e capitoli. Quest'ultimo, inoltre, presenta prima dell'*Etica Nicomachea* un *escerto* dalla traduzione degli *Economici* pseudo-aristotelici, eseguita sempre da Bruni (cfr. almeno Josef Soudek, *A Fifteenth-Century Humanistic bestseller: the manuscript diffusion of Leonardo Bruni's annotated Latin Version of The (Pseudo-) Aristotelian Economics*, in: *Philosophy and Humanism: Renaissance essays in honor of Paul Oskar Kristeller*, edited by Edward Patrick Mahoney, Leiden: Brill, 1976, p. 129-143, con bibliografia.

[20] Nel ms. D V 33 la versione degli *Economici* è preceduta dalla dedica di Bruni a Cosimo de' Medici.

MODUS

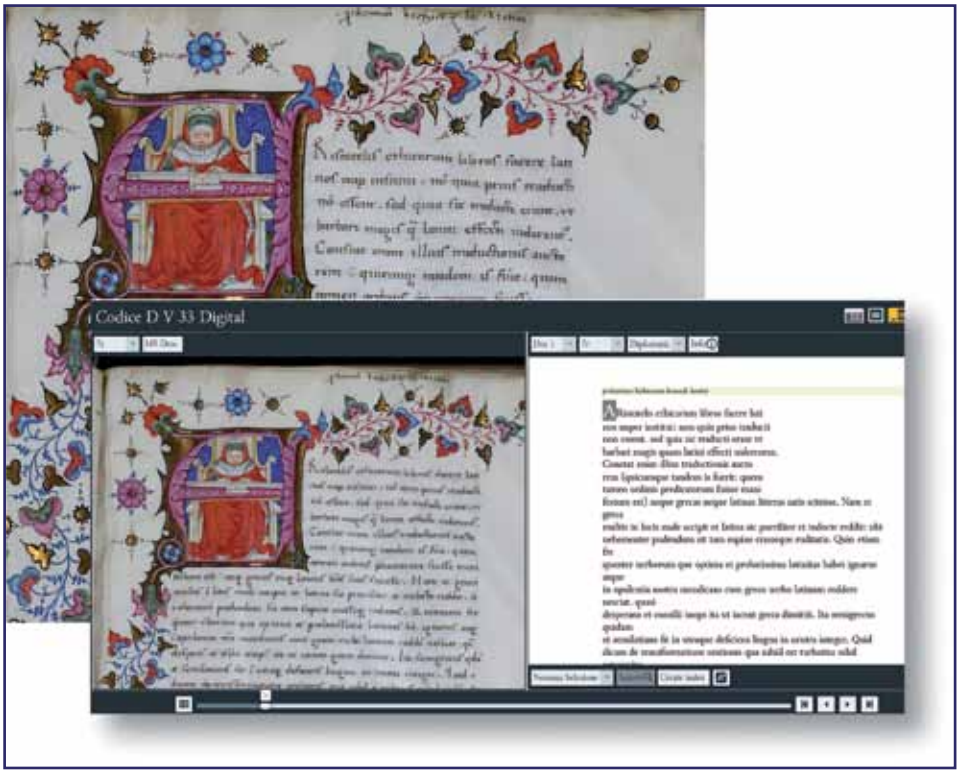

Figura 3. *L' edizione digitale del manoscritto D V 33 visualizzata attraverso il software EVT*

L'edizione di Plinio, limitata alle 31 carte miniate, è anch'essa di tipo diplomatico interpretativo ed è dotata di un *markup* descrittivo analogo a quello usato per BINARI, cui ho aggiunto delle annotazioni di tipo filologico: in particolare, ho registrato le differenze testuali rispetto all'edizione critica di Karl Mayhoff (1892-1909, disponibile online sul PHI Latin Texts[21] e rispetto al codice da cui è stato copiato l'esemplare torinese, cioè Milano, Biblioteca Ambrosiana, D 531 inf.[22]. Infine, ho provveduto a creare degli indici inserendo dei tag ai toponimi e ai nomi di persona menzionati nel testo.
Il differente grado di sviluppo della marcatura nei due progetti è stato condizionato principalmente dalla letteratura a disposizione. Per la *Naturalis Historia* di Plinio potevo contare su diversi studi (penso, in particolare, a quello di Michael Reeve sulla storia della tradizione del testo, con un'attenzione anche per i testimoni *recentiores*)[23] e, soprattutto, su un'edizione critica, mentre le versioni di Bruni e

[21] <https://catalog.perseus.org/catalog/urn:cts:latinLit:phi0978.phi001.perseus-lat1>. Oltre al testo critico, nell'edizione digitale ho indicato anche il link a cui trovare una traduzione (in inglese) del testo pliniano: Pliny the Elder, *The Natural History*, ed. by J. Bostock, H.T. Riley, London: Taylor and Francis, 1855, <http://data.perseus.org/texts/urn:cts:latinLit:phi0978.phi001.perseus-eng1>.

[22] Consultabile in versione digitalizzata al seguente link: <https://digitallibrary.unicatt.it/veneranda/0b02da8280051bee>.

[23] Cf. M. Reeve, *The editing of Pliny's Natural History*, cit., p. 107-179.

Argiropulo sono ancora inedite, così come pochi e parziali sono ad oggi gli studi sulla loro trasmissione: lacune, queste, che non hanno consentito l'inserimento di annotazioni di tipo filologico. Inoltre, il *markup* più strutturato per il testo pliniano è stato favorito dal numero contenuto delle carte edite (31 contro i più di 100 *folia* degli altri codici). Tuttavia, il grado di maturazione raggiunto dal progetto non è da ritenersi definitivo: le caratteristiche di EVT, sostenibilità nel tempo e modularità, e, più in generale, del linguaggio XML-TEI, lasciano la porta aperta a future implementazioni di queste edizioni, in parallelo con il progredire degli studi (ad esempio il testo critico della versione di Argiropulo da me stabilito potrà fornire dati utili a sviluppare il *markup* almeno per il codice E III 25).

Dopo aver dato conto degli strumenti utilizzati, delle modalità e dei criteri editoriali seguiti nell'allestimento delle edizioni, occorre soffermarsi su come queste si presentino concretamente al fruitore. Innanzitutto, ad esse si accede attraverso il sito della BNUTO e, in particolare, nella sezione "Manoscritti in EVT" all'interno della Biblioteca digitale[24]; cliccando sulla segnatura di ciascun manoscritto si arriva alla rispettiva edizione in EVT (Fig. 4): la videata principale è bipartita[25], con a destra la trascrizione del testo della carta di cui, a sinistra, è presente il *facsimile* digitale[26], ingrandibile a piacimento; in basso le frecce consentono di spostarsi all'interno dell'edizione, navigabile anche selezionando il numero del folio desiderato negli appositi riquadri della barra in alto (per Plinio, il riferimento è alla miniatura). Le informazioni relative all'edizione e alla fonte (ovvero il manoscritto) si trovano nella sezione "Paratesto e indici", raggiungibile attraverso un'apposita icona (la prima da sinistra del gruppo di quattro, posto in alto a destra); al suo interno sono raccolti i dettagli sul progetto ("Informazioni sul progetto"), quali i nominativi dei responsabili, le informazioni relative alla pubblicazione; la descrizione del codice edito ("Descrizione della fonte")[27]; la bibliografia ("Bibliografia") e gli indici ("Named Entities", per i toponimi e antroponimi; "Indice dei contenuti"). Quanto alle note riguardanti la trascrizione (i criteri seguiti, la legenda dei colori usati nel *markup* e, per Plinio, le informazioni circa le annotazioni filologiche), ad esse si accede attraverso l'icona "Info", nella pagina principale, in alto a destra.

[24] https://bnuto.cultura.gov.it/biblioteca-digitale/manoscritti-in-evt/.

[25] Questo layout è modificabile dall'utente: all'estrema destra, in alto, trova le icone "Testo Testo" (la prima a partire da destra) e "Testo di lettura" (la terza; la seconda icona, "Immagine Testo" è quella che appare di default), le quali consentono di mostrare a video solo il testo.

[26] Le miniature di tutti i facsimili sono visibili complessivamente cliccando sull'apposita icona ("Miniature") posta in basso a sinistra, accanto alla barra che consente lo spostamento all'interno dell'edizione.

[27] Essa è raggiungibile anche dalla schermata principale, cliccando sull'icona "MS Desc", in alto a sinistra.

MODUS

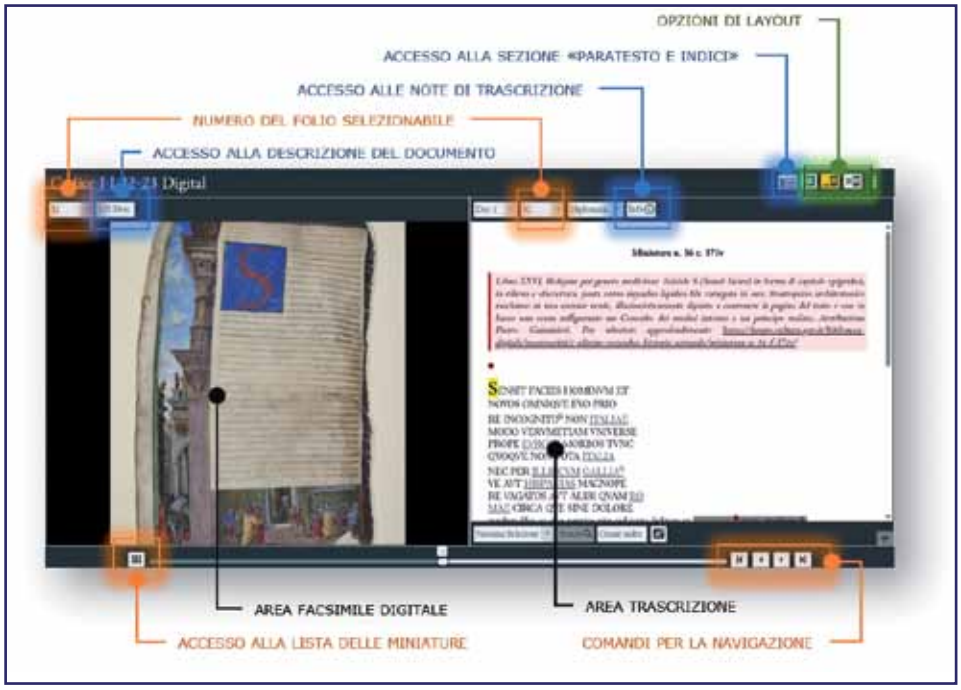

Figura 4. *Legenda dei principali comandi di visualizzazione del software EVT*

Nel caso di Plinio, la schermata principale ospita anche una breve descrizione dell'apparato figurativo ad opera della prof.ssa Federica Toniolo, ordinaria di Storia dell'arte medievale presso l'Università degli studi di Padova, argomento di cui, però, parleremo più diffusamente nel prossimo paragrafo.
Da ultimo, una nota sui facsimili digitali impiegati: se, da un lato, avere a disposizione immagini di qualità elevata facilita la fase di trascrizione, dall'altro nella pratica le rende del tutto inutilizzabili all'interno delle edizioni digitali, dato il loro peso in termini di occupazione di memoria. Perciò, come già evidenziato in precedenza, si è resa necessaria un'operazione di compressione delle immagini stesse ad un livello molto alto per renderle compatibili con il visualizzatore: a titolo di esempio, le immagini originali del Plinio, ciascuna del peso tra gli 80 ed i 90 megabyte, sono state ridotte a file dell'ordine dei 400-500 kilobyte. Tutto questo ovviamente incide sulla loro qualità effettiva, ma non su quella percepibile, in quanto ottimizzata per una riproduzione su schermo fino ai livelli di ingrandimento richiesti dal software di visualizzazione.

Digitalizzazione e interdisciplinarità

Si è accennato a come il codice di Plinio sia assai rilevante per la storia dell'arte: infatti, il suo apparato decorativo e illustrativo costituisce uno dei vertici della miniatura italiana del Rinascimento, in quanto specchio della qualità raggiunta dalla cultura visiva mantovana negli anni in cui, dopo Pisanello, Andrea Mantegna era pittore alla corte

dei Gonzaga. Un'edizione del manoscritto, dunque, non poteva prescindere dal rendere conto di tale eccezionalità, vista anche l'occasione di sfruttare le possibilità offerte dal digitale di superare le limitazioni del cartaceo in materia di quantità di informazioni e di fruizione delle immagini. Si è scelto, quindi, di corredare l'edizione diplomatica di una descrizione delle miniature e, a tal scopo, è stata attivata una collaborazione con la già citata prof.ssa Toniolo.

La docente ha redatto un'introduzione generale in cui descrive complessivamente l'apparato decorativo, contestualizzandolo sul piano storico e geografico e all'interno della produzione artistica coeva, e affronta la complessa questione dell'attribuzione delle miniature, fornendo anche preziose indicazioni bibliografiche. La prof.ssa Toniolo, inoltre, si è occupata della descrizione di ciascuna carta miniata: come preannunciato, una versione più breve è presente sulla pagina EVT, in un apposito riquadro anteposto alla trascrizione, ma ne è disponibile anche una più dettagliata su una pagina web dedicata, raggiungibile dalla stessa edizione in EVT (il link è indicato al termine della descrizione compendiata). Tale sezione del sito[28] è tripartita: la videata principale riporta la storia del codice pliniano ("La storia"), il testo introduttivo dell'apparato illustrativo ("L'opera") e l'elenco di tutte le 31 miniature ("Le miniature")[29]; cliccando su ciascuna voce della lista si accede a una pagina con l'immagine della carta decorata, su cui sono indicate tre zone (la lettera capitale, il fregio laterale e il tondo decorato nel margine inferiore); spostando il puntatore su ciascuna di queste aree si apre un riquadro con la descrizione relativa a quello specifico dettaglio della decorazione con un ingrandimento dello stesso.

Conclusioni

Volendo fare un bilancio di questa esperienza, i progetti Il Plinio di Mantegna e BINARI hanno messo in luce le potenzialità del digitale non solo nella conservazione e nella salvaguardia del patrimonio manoscritto, ma anche nello studio di esso da parte di un pubblico di esperti in diversi settori disciplinari. Ciò è stato reso possibile grazie all'uso di determinati strumenti informatici che hanno consentito di corredare l'immagine del manufatto con una serie di informazioni, selezionate e inserite sulla base di criteri scientifici, precedentemente fissati, di carattere filologico, storico e artistico. Inoltre, la dotazione delle edizioni di appositi sussidi, quali la trascrizione del testo (e/o eventualmente una sua traduzione) e la marcatura di toponimi e antroponimi, permetteranno di proporre e divulgare il risultato finale anche presso una più ampia cerchia di fruitori non specialisti, con l'intento di sensibilizzare l'opinione pubblica sulla ricchezza del patrimonio librario e, al contempo, sull'importanza del lavoro svolto dalle biblioteche per la sua salvaguardia.

Da questi progetti, inoltre, risulta evidente che, per realizzare un prodotto digitale corredato da contributi di natura differente per venire incontro a più esigenze, occorrono molteplici competenze anche in ambito informatico oltre che specialistico, per acquisire

[28] https://bnuto.cultura.gov.it/biblioteca-digitale/manoscritti/c-plinius-secundus-historia-naturalis.

[29] Prima dell'elenco è riportato il link al testo in EVT, di modo che le due sezioni dell'edizione siano comunicanti in entrambi i sensi.

MODUS

le quali ho seguito percorsi di formazione specifici nell'ambito delle *Digital Humanities*. Allo stesso modo, questa esperienza dimostra la necessità di avere contatti ed avviare collaborazioni tra istituzioni diverse (BNUTO, le università di Torino, Innsbruck e Padova, il centro DISH, la cooperativa Readcoop).
Un'altra caratteristica dei prodotti digitali creati è il loro essere potenzialmente sempre *in fieri*: nella realizzazione di questi progetti, infatti, si è avuta cura che possano trarre beneficio dal progredire degli studi, ovvero che siano sempre passibili di successive espansioni e aggiornamenti, anche grazie alla scelta delle soluzioni informatiche impiegate (XML, EVT); in particolare, si è badato che essi possano durare nel tempo, così da ridurre il rischio che non siano più supportati da tecnologie in costante evoluzione.
Da ultimo, riflettendo sulle opportunità offerte dai processi di digitalizzazione, i progetti Il Plinio di Mantegna e BINARI si pongono come pratici riferimenti per iniziative analoghe e come base per evoluzioni successive: ad esempio, una futura edizione critica digitale delle versioni latine dell'*Etica Nicomachea* potrà avvalersi, per la fase di collazione, dell'edizione dei testimoni D V 33, E III 25 e E IV 36; viceversa, la trascrizione di questi potrà essere arricchita con note filologiche e rimandi al testo critico.

Since 2022, the National University Library of Turin has been involved in research and promotion activities focusing on a number of valuable manuscripts belonging to its collections. Such activities are carried out in close collaboration with the Interdepartmental Research Centre for Digital Scholarship for the Humanities (DISH) at the University of Turin and they include the projects Il Plinio di Mantegna and BINARI: La Biblioteca Nazionale per Aristotele. Digitalizzazione ed edizione digitale delle traduzioni latine umanistico-rinascimentali dell'Etica Nicomachea di Aristotele nei codici della Biblioteca Nazionale di Torino, both funded by PNRR resources. In particular, BINARI provides a digital edition of the manuscripts containing Latin translations of Aristotle's Nicomachean Ethics *by Leonardo Bruni (ms. D V 33, E IV 36, G III 16, G IV 32 and H V 16) and John Argyropoulos (ms. E III 25), while the former project aims to offer a digital edition of the famous illuminated manuscript of Pliny the Elder's* Naturalis Historia *(ms. J.I.22-23), which testifies the reception of Andrea Mantegna's art. This article aims to describe both these projects, outlining their various phases in order to highlight information technology potential not only as regards the conservation and preservation of manuscripts heritage, but also their study by a specialists audience, promoting an interdisciplinary approach and the dissemination to a wider public.*

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

The watermark portal Bernstein-Memory of Paper and its Italian contributions

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00146

Emanuel Wenger
Austrian Academy of Science

Watermarks are studied, collected, and catalogued since more than 200 years. They can give important information about the date, origin, chronology, and authenticity of paper documents. The result of these studies is a huge amount of watermark catalogues and publications. Nowadays, watermark catalogues are published in digital form as online databases. The idea of the Bernstein project is the development of a portal which allows the simultaneous access to watermark databases. In 2025, 56 databases from 27 countries are connected to the project and these numbers are still growing. This article focuses on the important Italian contributions to the project.

1. Introduction

In the second half of the 13th century, a few important innovations in the paper production were introduced in Italy. These innovations were mainly implemented by papermakers in Fabriano who also introduced watermarks[1]. Italian paper became popular fast. It was traded all over Europe because of its high quality.

The interest in watermarks among historians comes primarily from the fact that watermarks can help dating undated paper based documents. This was already well known at the beginning of the 19th century. Dating is done by comparing watermarks from undated paper with dated documents carrying the same watermark. Hence, historians in many countries collected watermarks from dated material and published them in watermark catalogues.

Besides the dating, watermarks are frequently used for the investigation of provenance, reception, work genesis and material layers, authenticity and attribution, genre history, collection contexts, assignment of anonymous or source fragments and writing developments. They play a role in many scientific fields like musicology, art history, palaeography, history of literature and other disciplines. In many humanities projects involving paper, the recording of watermarks is often a necessary step.

One special use of watermarks should also be mentioned here, namely that of detecting forgeries. In printmaking in particular, the paper is often the only indicator of whether the print is a forgery or reprint, or a so-called original.

[1] Sylvia Rodgers Albro, *Fabriano: City of Medieval and Renaissance Papermaking*, New Castle: Oak Knoll Press, 2016.

MODUS

Among the first watermark researcher were the scholars Aurelio (1830-1902) and Augusto Zonghi (1840–1916) who studied paper history and watermarks from Fabriano. Augusto Zonghi collected 2.213 paper samples from Fabriano papers manufactured between 1267 and 1798 and he made a catalogue with 1.887 watermark tracings ordered by motifs and date. This collection of his tracings was printed not until 1953[2].
The literature about watermarks published for more than 200 years is large, widespread and, most of it, difficult to access. The International Association of Paper Historians (IPH) web page "Printed Watermark Repertories" contains a list of watermark catalogues[3]. Nowadays, watermarks are increasingly recorded digitaly and stored in online databases. This offers the opportunity to connect such material online. Hence, the main objective of the project was to develop a web portal which gives access to a variety of watermark information online.

2. The eContentPlus project Bernstein (Memory of Paper)

In 2005, the European Union (EU) published the call for the programme eContentPlus (2005-2008)[4], the successor of the eContent programme (2001-2004). The purpose of the eContentPlus programme was to stimulate the development of digital content and services in areas of public interest, to reduce the fragmentation of the European digital content market and to improve access and ease of use of geographical information, educational and scientific material and cultural content.
The project *Bernstein – Memory of Paper* – was accepted for financing with a budget of 3.2 million € by eContentPlus. The project started in September 2006 with a duration of 30 months. It was a cooperation between nine institutions from Austria, England, France, Germany, Italy, and the Netherlands[5].
The objective of the Bernstein project was to create a multilingual integrated European digital environment for the expertise and history of paper with a focus on watermarks. The project should interlink all existing European watermark databases and also provide access to a bibliography centered on the topic of papermaking. There were four European watermark databases in 2006:

– Piccard-Online (POL, Stuttgart, Germany, 92.000 watermarks)[6];
– Watermarks in Incunabula printed in the Low Countries (WILC, The Hague, Netherlands, 11.000 watermarks)[7];
– Watermarks of the Middle Ages (WZMA, Vienna, Austria, 9.000 watermarks)[8];
– Dutch Institute of Art history in Italy Watermark Database (NIKI, Florence, Italy, 2.500 watermarks)[9];

[2] Aurelio Zonghi — Augusto Zonghi — Andrea Federico Gasparinetti, *Zonghi's watermarks,* (Monumenta chartae papyraceae historiam illustrantia; 3), Hilversum: The Paper Publications Society, 1953.
[3] https://www.paperhistory.org/Watermark-catalogues.
[4] https://eur-lex.europa.eu/EN/legal-content/summary/econtentplus-2005-2008.html.
[5] https://memoryofpaper.org/.
[6] Now part of Wasserzeichen Informationssystems (WZIS): <https://www.wasserzeichen-online.de>.
[7] https://watermark.kb.nl/page/index/title/Introduction.
[8] https://www.wzma.at/.
[9] http://www.wm-portal.net/niki/index.php.

The official project ended in February 2009 with a final report and the webpage[10] giving access to four databases as the main result[11]. Additionally, a bibliography with more than 31.000 bibliographic entries became searchable.
Italian scholars and institutions were involved in the project from the beginning with the involvement of the Netherlands Interuniversity Art Historical Institute (NIKI) in Florence who contributed with an international database of watermarks and paper used for prints and drawings. It was also the scholar Ezio Ornato who gave important ideas concerning watermark catalogues, statistics and geographical distributions of watermarks and the two scholars at the Istituto centrale per la patologia degli archivi e del libro (ICPAL) in Rome, Paola Munafò and Viviana Nicoletti who organized an exhibition for the project and worked specifically for the watermark terminology used in the project, and the Italian translations of the Bernstein websites.

3. Languages

One important requirement for most EU projects is the multilingualism. Six languages were chosen which had developed specific terminologies for watermark descriptions. These were English, French, German, Italian, Russian, and Spanish. Sources for the terminologies were significant watermark catalogues in these languages, like the watermark catalogues of the *Monumenta Chartae Papyricae* series for English, Briquet for French[12], Piccard for German[13], Zonghi for Italian[14], Likhachev for Russian[15], and Basanta Campos for Spanish[16], to mention only a few. Four additional languages were added a couple of years later, after the end of the project. These were Hungarian, Portuguese, Greek, and Dutch. Hence, the Bernstein portal supports 10 languages in 2025.

Figure 1. *Italian home page of the Bernstein portal*

[10] https://www.memoryofpaper.eu.
[11] Wenger 2021.
[12] Briquet 1907.
[13] Piccard 1961-1997.
[14] Zonghi 1881.
[15] Simmons et al. 1994.
[16] Basanta Campos 1996-2005.

Two areas are language dependent. One is the language of the webpages, the so-called user interface. Such translations can nowadays be done by web browsers with automatic translation features.
The more difficult and important issue is the watermark terminology. An automatic translation between languages for watermarks cannot be used because the watermark descriptions are derived from the various watermark terminology traditions. One example is the watermark type *bull's head* which is in German *Ochsenkopf* (ox head) and not *Stierkopf* (bull's head). Hence, the Bernstein project had to organize its own translation of watermark terms with a dictionary adapted to the watermark terminology[17]. Paola Munafò and Viviana Nicoletti worked on the Italian translations.

4. Databases

The main part of the project is the portal to watermark databases. The website allows the multilingual search in all the databases connected with the portal. Search items are the motif, place of use, depository, date, width and height, paper mill, paper maker, distance between chain lines, and some more. 56 watermark databases were connected to the portal in May 2025.

4.1. Italian watermark databases

Four databases consist of watermark collections hosted in Italy. Additionally, Italian watermarks can be found in many non-Italian watermark collections. Catalogues like Briquet, Piccard and many other contain numerous watermarks from Italian paper mills.

4.1.1. NIKI

The watermark database[18] of the Dutch Institute of Art history in Italy (NIKI) in Florence was among the first online watermark databases. It was part of the project since the beginning. The watermarks dated mainly from the 17th century in contrast to the other three Italian databases with contents from the Middle Ages until the 16th century. They were taken from drawings of mainly Dutch painters working in Italy. The recording of the watermarks was done by transparency photography with Photoshop postprocessing. This was done in cooperation with Bert W. Meijer and Georg Dietz.

4.1.2. CCI

Corpus Chartarum Italicarum[19] (CCI) is the biggest database among the Italian collections. It contains paper sheets, watermarked or not. The corpus of data was initially collected by the Istituto centrale di Patologia del Libro (ICPL) in the early 1940s and increased in the following decades. The Corpus illustrates Italian paper production and its use as a support for written memory from the 13th to the 20th century. It represents the

[17] https://www.memoryofpaper.eu/products/watermark_terms_v11.0_en.pdf.
[18] http://www.wm-portal.net/niki/index.php.
[19] https://cci-icpal.cultura.gov.it/.

most important collection of Italian paper sheets in terms of quantity and importance; it consists of various types of Western paper sheets produced starting from the mid-13th century when, in Fabriano, papermakers made significant changes to the structure and appearance of the sheets produced in the Arab-Hispanic context. The database contains a total of 5.362 watermarks. The contact persons for the Bernstein portal were Paola Munafò and Viviana Nicoletti and, more recently, Alessandra Fucini.

4.1.3. CCF

The Corpus Chartarum Fabriano[20] (CCF) is a catalogue of digital images of both watermarked and non-watermarked papers, manufactured in Fabriano over an uninterrupted span of eight centuries**,** from the late 13th century to the present day. The project involves the identification, cataloging, and digitization of the paper sheets; this data is then uploaded to a cutting-edge database designed to meet the requirements of the relevant scientific community. Livia Faccioni from the Fedrigoni Fabriano Foundation (FFF) cooperates with the Bernstein portal and makes the integration of CCF into the Bernstein portal possible.

4.1.4 Mazzoldi

The database Mazzoldi contains the watermarks which were collected by Leonardo Mazzoldi and published in two volumes. It contains 1.036 watermarks from the area of Brescia close to the Garda Lake[21].

5. Catalogues, exhibitions, and conferences

5.1. Catalogues

The Bernstein project was only possible in close cooperation with the Landesarchiv Baden-Württemberg, Hauptstaatsarchiv Stuttgart. The entire Piccard Index collection with more than 92.000 watermarks was digitized and put online as Piccard-Online. The quantity and quality of this database was essential for the acceptance of the project by the EU in the frame of the program eContentPlus.

In connection with Piccard-Online, the Hauptstaatsarchiv Stuttgart under the guidance of Peter Rückert designed and organized a German and Italian exhibition. The title of the exhibition was "Ochsenkopf and Meerjungfrau – Wasserzeichen des Mittelalters", respectively "Testa di bue e sirena – la memoria della carta e delle filigrane dal medioevo al seicento"[22]. These exhibitions were adopted by the Bernstein project as part of the dissemination activities. The exhibitions were held in Italy, Germany, and Austria. The exhibitions were accompanied by a booklet in German (2006) and an Italian extended edition in 2007. The concept of the Italian exhibition and the Italian catalogue were mainly done by Paola Munafò, Viviana Nicoletti, Franco Mariani, Ezio Ornato, and

[20] https://www.fondazionefedrigoni.it/it/il-progetto.

[21] *Filigrane di cartiere bresciane*, a cura di L. Mazzoldi, Brescia: Ateneo di scienze lettere ed arti, 1990-1991.

[22] https://memoryofpaper.org/twiki/pub/Main/ProjectExhibitions/bernstein_2007_exhibition_catalog_it.pdf.

MODUS

Giorgio Pellegrini. A third extended edition in English and German was published 2009. These booklets are self-contained books with 128 pages about the history of paper and watermarks from the Middle Ages to the modern period[23]. An extended Spanish version with the title "Cabeza de Buey y Sirena" was published in 2011[24] and is to this day the last edition.

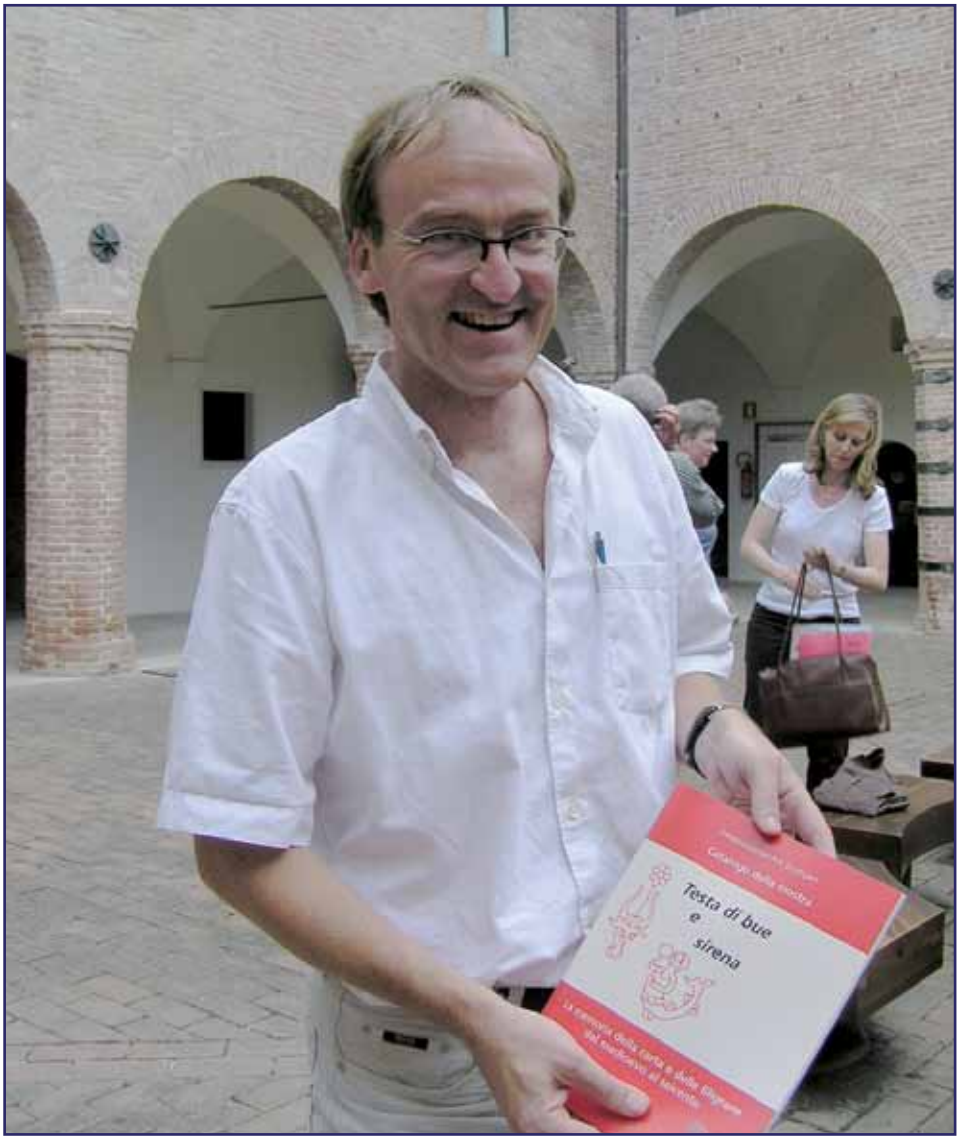

Figure 2. *Presentation of "Testa di bue e sirena" in Fabriano by Peter Rückert*

[23] https://memoryofpaper.org/twiki/pub/Main/ProjectExhibitions/bernstein_2009_book_en.pdf.

[24] Rückert — Perez Garcia — Wenger 2011, <https://memoryofpaper.org/twiki/pub/Main/ProjectExhibitions/bernstein_2011_book_es.pdf>.

MODUS

5.2. Exhibitions

The series of exhibitions started in Stuttgart 2006 at the Archives of the State of Baden-Württemberg and continued 2007 in Vienna in the Museum of the Schottenstift.
Alread in July 2007, the first Italian exhibition (9 July – 24 August) took place at the Paper and Watermark Museum in Fabriano, the city of paper and watermarks. This exhibition was made possible by Franco Mariani and Giorgio Pellegrini. The series continued in September 2007 (8 September – 10 October) at the ICPAL organized by Paola Munafò and Viviana Nicoletti. This institute became one of the major supporters of the project. Among others, it contributed the database CCI.

Figure 3. *Exhibition at ICPAL, 2007*

The third Italian exhibition was opened at the Castello Sforzesco, Sala del Tesoro della Trivulziana in Milano with the title "Una firma nell'acqua - La memoria della carta e delle filigrane dal Medioevo all'età moderna" (28 October - 31 December 2008). Franca Porticelli was the local organizers of the exhibition. This exhibition was also the first contact with Timoty Leonardi. He became the most important supporter of the series of exhibition and the conferences. A big thanks goes to him.
The first exhibition which Timoty Leonardi organized together with Isabella Fiorentini was at the Biblioteca Nazionale di Torino with the title "Testa di bue e sirena. Cartiere e

filigrane piemontesi: prospettive di ricerca" (30 April – 27 June 2009). This exhibition was accompanied by an additional catalogue edited by Timoty Leonardi[25]. 2011 followed the exhibition (6 May-10 July) in Vercelli at the Museo del Tesoro del Duomo, which was the working place of Timoty Leonardi.

Figure 4. *Exhibition in Vercelli, 2011*

Finally, the sixth and last exhibition took place in the wonderful city of Varallo at the Biblioteca Civica di Varallo in 2012 (15 June – 1 July). The Italian series of exhibitions became the most successful, wonderful, and important, one thanks to Timoty Leonardi.

Figure 5. *Exhibition opening in Varallo, 2012*

[25] https://memoryofpaper.org/twiki/pub/Main/ProjectExhibitions/bibnattorino_2009_exhibition_catalog_it.pdf.

MODUS

5.3. Conferences

The Bernstein project held yearly meetings and work package meetings during the running time from September 2006 until February 2009. The first yearly meeting took place in Fabriano at the museum of paper and watermarks.

Figure 6. *First annual Bernstein meeting in Fabriano, July 2007*

The series of meetings was discontinued in February 2009 with the official end of the EU project. But after a couple of years, the former team of the project wanted to meet again. Hence, a series of international conferences on watermarks in digital collections in connection with the Bernstein portal was started in 2013 in The Hague.

Figure 7. *Conference at Fedrigoni Fabriano Foundation, 2022*

MODUS

The series got the name: *International conferences on watermarks in digital collections*. It continued in Valencia (Spain, 2014), Santa Maria da Feira (Portugal, 2016), Vienna (Austria, 2017), and Cork (Ireland, 2019). After a break due to the Corona epidemy, the conference was hosted by the Fedrigoni Fabriano Foundation in Fabriano in 2022. Livia Faccioni had organized a wonderful conference which was documented totally on YouTube videos[26].

Already one year later, this means 2023, Timoty Leonardi organized the sixth conference in Verona at the Fondazione Biblioteca Capitolare di Verona with an interesting excursion to the paper museum in Toscolano. In 2025, the conference is held in Reykjavik.

6. Conclusion and Acknowledgements

The Bernstein portal became one of the most important tools for the identification of watermarks. The cooperation with the Italian watermark experts was an important part and enrichment of the project. I would like to thank all Italian colleagues for their wonderful work and their commitment to the Bernstein project and I hope for further cooperation. The project is still going on and new data is added. Further Italian watermark collections which are not connected with the Bernstein portal are invited to cooperate with the Bernstein project.

Da oltre 200 anni le filigrane vengono studiate, analizzate, raccolte e catalogate. Esse possono fornire informazioni importanti sulla data, l'origine, la cronologia e l'autenticità dei documenti cartacei. Il risultato di questi studi è un'enorme quantità di cataloghi e pubblicazioni sulle filigrane. Oggi i cataloghi delle filigrane sono pubblicati in formato digitale come banche dati online. Alla base del progetto Bernstein vi è l'idea dello sviluppo di un portale che consenta l'accesso simultaneo alle banche dati del settore. Ad oggi, 56 banche dati di 27 paesi sono collegate al progetto e questi numeri sono in crescita. Questo articolo si concentra sugli importanti contributi italiani al progetto.

[26] https://www.youtube.com/playlist?list=PLZLyD4fwo5L-fqKuLSotSiZ7y9ngAgupd.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

References

Albro 2016 — Sylvia Rodgers Albro. *Fabriano: City of Medieval and Renaissance Papermaking*. New Castle: Oak Knoll Press, 2016.

Briquet 1907 — Charles-Moïse Briquet. *Les filigranes: dictionnaire historique des marques du papier dès leur apparition vers 1282 jusqu'en 1600*. (4 volumes, first edition). Paris [etc.]: Picard [etc.], 1907.

Leonardi 2009 — Timoty Leonardi. *Cartiere e filigrane piemontesi: prospettive di ricerca: la memoria della carta e delle filigrane dal Medioevo al Seicento*. Torino: Biblioteca Nazionale Universitaria, 2009.

Mazzoldi 1990-1991 — *Filigrane di cartiere bresciane,* a cura di L. Mazzoldi. (2 volumi). Brescia: Ateneo di scienze lettere ed arti, 1990-1991.

Piccard 1961-1997 — Gerhard Piccard. *Die Wasserzeichenkartei Piccard im Hauptstaatsarchiv Stuttgart*. (Findbücher I-XVII). Stuttgart: W. Kohlhammer, 1961-1997.

Rückert 2007 — *La memoria della carta e delle filigrane dal Medioevo al Seicento: testa di bue e sirena*, a cura di P. Rückert. Stuttgart: Landesarchiv Baden-Württemberg, Hauptstaatsarchiv, 2007.

Rückert — Hodecek — Wenger 2009 — *The history of paper and watermarks from the Middle Ages to the Modern Period: bull's head and mermaid*, ed. by P. Rückert, S. Hodecek, E. Wenger. Stuttgart; Vienna: Bernstein Project, 2009.

Rückert — Perez Garcia — Wenger 2011 — *Cabeza de buey y sirena: la historia del papel y las filigranas desde el medievo hasta la modernidad*, ed. P. Rückert, C. Perez Garcia, E. Wenger. Stuttgart: Hauptstaatsarchiv, 2011.

Scianna 2018 — Nicolangelo Scianna. *Watermarked paper from archives in Ravenna* (1287-1693). (2 vol.). Turnhout: Brepols, 2018.

Simmons et al. 1994 — *Likhachev's watermarks: an english-language version*, ed. by J. S. G. Simmons, B. J. van Ginneken de Kasteele. (Monumenta Chartae Papyraceae; 15). Amsterdam: The Paper Publications Society, 1994.

Wenger 2021 — Emanuel Wenger. *The Bernstein Project. The Memory of Paper: a Multilingual Online Portal of Databases*. «IPH Congress Book», 23 (2021), p. 260-269.

Zonghi 1881 — Aurelio Zonghi. *Le Marche principali delle carte Fabrianesi dal 1293 al 1599*. Fabriano: Tipografia Gentile, 1881.

MODUS

Zonghi 2020 — Augusto Zonghi. *I segni delle antiche cartiere fabrianesi*, a cura di Livia Faggioni. Fabriano: Fondazione Gianfranco Fedrigoni, 2020. (First printed edition of the handwritten album drafted 1884 by Augusto Zonghi, with notes to the album in English and Italian).

Zonghi — Zonghi — Gasparinetti 1953 — Aurelio Zonghi — Augusto Zonghi — Andrea Federico Gasparinetti. *Zonghi's watermarks*. (Monumenta chartae papyraceae historiam illustrantia; 3). Hilversum: The Paper Publications Society, 1953.

Transkribus e la trascrizione automatica di manoscritti. Un caso studio sulle carte di Vittorio Alfieri

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00147

Sara Gallegati
Università degli studi di Macerata

Lo studio esplora l'uso della piattaforma Transkribus per la trascrizione automatica di manoscritti, utilizzando i testi inclusi nel manoscritto Alfieri 6 (conservato presso la Biblioteca Medicea Laurenziana di Firenze). Dopo una prima fase di test con il modello generico Italian Handwriting M1, i cui risultati si sono dimostrati insoddisfacenti, è stato creato un modello personalizzato a partire dal modello M1, che ha ottenuto un CER del 13.40%. Un secondo addestramento, ampliando il set di ground truth, ha ridotto il CER al 6.9%, con trascrizioni più accurate. Nonostante i progressi, permangono alcune difficoltà nella distinzione tra lettere simili; ma lo studio conferma l'efficacia di Transkribus per la trascrizione di testi manoscritti, suggerendo l'importanza di un numero adeguato di dati per migliorare la qualità della trascrizione automatica.

Introduzione

Transkribus è una piattaforma basata sull'intelligenza artificiale e sulle tecnologie di Handwritten Text Recognition (HTR); consente alle macchine di riconoscere testi manoscritti e a stampa, convertendo le immagini in trascrizioni codificate, ricercabili ed esportabili in vari formati.

L'idea alla base della piattaforma è che «the collaboration between (1) archives/libraries, (2) humanities scholars, (3) computer scientists/technology providers and (4) the public (crowd users, volunteers) is key for the success of an innovative platform which aims at substantially improving access to historical documents on the long term»[1].

Uno dei maggiori vantaggi di Transkribus è il suo essere *user-friendly*: gli utenti possono gestire l'intero processo di digitalizzazione senza bisogno di una formazione informatica specifica, possono inoltre creare i propri modelli di riconoscimento di scrittura a mano o addestrarne di già esistenti[2].

[1] Sebastian Colutto — Philip Kahle — Günter Hackl — Günter Mühlberger, *Transkribus. A Platform for Automated Text Recognition and Searching of Historical Documents*, «IEEE Computer Society» (2019), p. 463-466: 463. DOI: 10.1109/eScience.2019.00060.

[2] I modelli pubblici disponibili (allo stato attuale 140, <https://readcoop.eu/transkribus/public-models/>) consentono agli utenti di far analizzare i documenti manoscritti e a stampa senza bisogno di addestramento personalizzato.

MODUS

I dati e i documenti caricati dagli utenti su Transkribus sono privati, il che riduce le restrizioni imposte dal copyright[3], e vengono archiviati nella piattaforma in modalità centralizzata, in modo da costituire un ampio bacino che può essere utilizzato per tutti i processi di apprendimento automatico.
La piattaforma permette inoltre di condividere tra gli utenti collezioni, documenti e prototipi di HTR per l'addestramento di modelli più ampi: «users of Transkribus are therefore not only consumers but also producers of their own model(s) and data which can then also be re-used by and shared with other users»[4].
Ad oggi, le tecnologie di riconoscimento di testi scritti a mano (HTR) hanno raggiunto un elevato grado di maturità e sono in grado di produrre trascrizioni precise dalle immagini di manoscritti storici. Grazie all'intelligenza artificiale e all'applicazione delle reti neurali profonde, infatti, le tecniche di HTR hanno registrato notevoli progressi nell'ultimo decennio, e vengono sempre più impiegate da biblioteche e archivi, per accelerare la trascrizione di fonti primarie e facilitare la ricerca e l'analisi di testi storici[5].
Questo strumento è stato applicato per ottenere le trascrizioni del materiale del *Panegirico di Plinio a Trajano* contenuto in Alfieri 6, il manoscritto conservato presso la Biblioteca Medicea Laurenziana di Firenze (Fondo Alfieri, [Cat. Sala Studio 6][6]). Trattasi della più antica attestazione del testo a noi giunta: l'opera fu ideata e stesa da Vittorio Alfieri[7] a Pisa tra l'inverno e l'estate del 1785, poi rivista durante il soggiorno a Martinsbourg, nel 1786, quindi stampata una prima volta a Parigi nell'aprile del 1787 per i tipi di François-Denis Pierres, e nuovamente, dopo un'ulteriore revisione, nell'autunno del 1789, da Didot, questa volta assieme a due altre opere, l'ode *Parigi sbastigliato* e la favoletta *Le mosche e l'Api*. Alfieri 6 ospita – oltre al *Panegirico* alle cc. 29*r*-38*r* – i materiali relativi alla *Tirannide*, recante ancora il titolo *Del Tiranno e Della Tirannide*

[3] «All documents uploaded to Transkribus are private by default. They are stored on the servers of READ-COOP SCE (i.e. the company that develops and maintains the software). The servers are all located in Innsbruck, Austria, in a GDPR-compliant manner, and the data may be processed according to the *terms & conditions* on the READ-COOP SCE website» Cfr. *Transkribus Help Center*, <https://help.transkribus.org/uploading-files-to-transkribus>.

[4] S. Colutto *et al*., *Transkribus. A Platform for Automated Text Recognition and Searching of Historical Documents*, cit., p. 463.

[5] Joe Nockels — Paul Gooding — Sarah Ames — Melissa Terras, *Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research*, «Archival Science», 22 (2022), p. 367-392: 368.

[6] Sul fondo dell'autore conservato alla Laurenziana cfr. Simone Casini, *Il fondo di manoscritti alfieriani della Laurenziana. Appunti per una storia interna*, in: *Alfieri a Siena e dintorni*. *Omaggio a Lovanio Rossi*. *Atti della Giornata di Studi, Colle di Val d'Elsa (22 settembre 2001)*, a cura di A. Fabrizi, Roma: Domograf, 2007, p. 175-193.

[7] Per una panoramica su Alfieri, l'opera e il relativo contesto storico-culturale, si vedano almeno Vittore Branca, *Sbastigliamenti alfieriani fra miti solari e fede palingenetica, delirio pindarico e autobiografia poetica*, «Italica» 68 (1991), n. 4, p. 401-418; Guido Santato, *Le Mosche sul Panegirico: Alfieri "sbastigliato"*, «Lettere italiane», 46 (1992), n. 1, p. 57-92; Giuseppe Rando, *Alfieri europeo: le "sacrosante leggi". Scritti politici e morali – Tragedie – Commedie*, Soveria Mannelli: Rubbettino, 2007; *Quand Alfieri écrivait en français: Alfieri et la culture française*, a cura di C. Del Vento, G. Santato, Paris: Bibliothèque Mazarine, 2003; Christian Del Vento, *Il Principe e il Panegirico. Alfieri tra Machiavelli e De Lolme*, «Seicento & Settecento», 1 (2006), p. 149-170; Laura Sannia Nowé, *Una* institutio principis *moderna: il* Panegirico di Plinio a Trajano *di Vittorio Alfieri*, in: *Studi dedicati a Gennaro Barbarisi*, a cura di C. Berra, M. Mari, Milano: Cuem 2007, p. 489-526.

(cc. 3*r*-26*r*); il dialogo *La Virtù Sconosciuta* (cc. 41*r*-50*r*); il trattato *Del Principe e Delle Lettere* (cc. 51*r*-89*r*); le *Prose diverse pel Misogallo. Prose cinque* (cc. 91*r*-112*r*). L'opera nasce, secondo quanto narrato nell'autobiografia[8], dalla lettura del panegirico che Plinio il Giovane recitò all'imperatore Traiano in senato nel 100 d.C.; e si configura come un'orazione il cui nucleo centrale ruota attorno alla richiesta di Plinio rivolta a Traiano di deporre volontariamente il potere assoluto e ristabilire la libertà repubblicana a Roma.
I vantaggi offerti dalla piattaforma Transkribus per lo studio di manoscritti sono molteplici: dopo una prima fase di preparazione del materiale si riducono drasticamente i tempi di trascrizione; è possibile aggiungere dei metadati ai testi così da arricchire le informazioni all'interno dei file; i documenti sono inoltre esportabili in diversi formati, tra cui in xml, aspetto fondamentale per permettere una successiva codifica secondo le linee guida TEI.

Flusso di lavoro per il ms. Alfieri 6

Il flusso di lavoro per la trascrizione del manoscritto si è composto di tre fasi: 1) creazione di una collezione e caricamento del materiale; 2) analisi del layout, selezione e verifica di un modello, *text recognition*; 3) operazioni di postproduzione, in particolare controllo e correzione delle trascrizioni ottenute.
La scelta di trascrivere Alfieri 6 con Transkribus non è dovuta tanto alla complessità grafica del manoscritto – la scrittura di Vittorio Alfieri, infatti, risulta piuttosto regolare e leggibile, al punto da non richiedere competenze paleografiche specialistiche – quanto alla possibilità di sfruttare questo caso come punto di partenza per la creazione di un modello HTR dedicato e altamente specifico. L'obiettivo è quello di sviluppare un modello in grado di riconoscere con precisione la mano di Alfieri, che potrà essere progressivamente raffinato e potenziato con l'inserimento di nuovo materiale autografo. In questo modo, il lavoro non solo ottimizza la trascrizione di Alfieri 6, ma pone le basi per una risorsa condivisibile e riutilizzabile da altri studiosi, favorendo una più ampia diffusione e valorizzazione digitale del *corpus* alfieriano.
Dopo la creazione delle collezioni sono stati caricati i file contenenti le carte del manoscritto. La piattaforma permette il caricamento di file in diversi formati: sono accettati sia file di immagini nei formati jpeg e png, con una dimensione massima di 10 MB ciascuno, che file in pdf fino a 3.000 pagine e con una dimensione massima di 200 MB. Per Alfieri 6 è stato caricato un unico file pdf, ogni pagina del documento è stata poi estratta e caricata automaticamente come singola. La buona qualità delle immagini non ha reso necessario alcun trattamento prima della trascrizione automatica.

[8] «In queste semiletture avea scorse le lettere di Plinio il Minore [...]. Finite l'epistole, impresi di leggere il Panegirico a Traiano, opera che mi era nota per fama, ma di cui non avea mai letta parola. Inoltratomi per alcune pagine, e non vi ritrovando quell'uomo stesso dell'epistole, e molto meno un amico di Tacito [...], io sentii nel mio intimo un certo tal moto d'indegnazione; e tosto, [...] impugnata con ira la penna, ad alta voce gridando dissi a me stesso: "Plinio mio, se eri tu davvero e l'amico, e l'emulo, e l'ammiratore di Tacito, ecco come avresti dovuto parlare a Traiano". E senza più aspettar, né riflettere, scrissi d'impeto, quasi forsennato, così come la penna buttava, circa quattro gran pagine del mio minutissimo scritto» (Vittorio Alfieri, *Vita*, in: *Opere*, Introduzione e scelta di Mario Fubini, testo e commento a cura di Arnaldo Di Benedetto, Milano-Napoli: Ricciardi, 1977, v. I, p. 251).

È inoltre possibile editare le collezioni create, inserendo metadati descrittivi sulle base delle regole xml TEI[9]. Le informazioni aggiunte possono essere poi esportate, assieme ai contenuti della collezione, in vari formati[10], tra cui xml codificato in TEI e ALTO (Analyzed Layout and Text Object[11]). La pagina esportata nel formato xml registra le informazioni relative alle caratteristiche dell'immagine, oltre che alla struttura del layout e al contenuto della pagina[12]; la pagina esportata in ALTO presenta uno schema XML che dettaglia i metadati tecnici per descrivere il layout e il contenuto delle risorse testuali fisiche, e viene spesso utilizzato in combinazione con METS.

Dopo aver creato ed editato la collezione, l'analisi si è concentrata sulle singole pagine. È stato quindi effettuato un censimento dei modelli presenti nella piattaforma per testi manoscritti in lingua italiana. I modelli proposti sono tre: *Transkribus Italian Handwriting M1*, *Italian Administrative Hands, 1550-1700* e *Ligorio 0.3 Pyl*. *Ligorio 0.3 Pyl* (CER[13] 2.7%), basato sulla grafia dell'antiquario italiano Pirro Ligorio (Napoli 1513 – Ferrara 1583)[14]. *Italian Administrative Hands, 1550-1700* (CER del 12.2%) è costituito da documenti in lingua italiana provenienti dagli archivi di Stato di Milano, Venezia, Firenze, Pisa e Genova: «the training set represents a spectrum of humanistic, italic and cursive hands characteristic of administrative records, employed by secretaries and newswriters»[15].

L'eccessiva specificità dei modelli appena descritti ha vincolato la scelta al modello *Transkribus Italian Handwriting M1* per un primo test di trascrizione. Il modello, che presenta un tasso di errore dei caratteri (CER) del 6.7%, è addestrato su testi di grafia italiana compresi tra il XVI e il XIX secolo. Tuttavia, «this is a generic model trained on a diverse dataset. Such models provide good results without the need for any extra training work. However, the best results can usually be achieved by training a special

[9] XML, acronimo di eXtensible Markup Language è un metalinguaggio creato e gestito dal World Wide Web Consortium (W3C); è una semplificazione e adattamento di SGML, da cui è nato nel 1998, e permette di definire la grammatica di diversi linguaggi specifici derivati. È un linguaggio di marcatura, ossia un linguaggio dotato di una semantica e di una sintassi specifiche in grado di mostrare all'utente tutte le annotazioni di caratterizzazione. I testi codificati in XML possono essere poi utilizzati sia per creare *output* orientati alla visualizzazione, sia come base di dati per effettuare delle elaborazioni automatiche. Nell'ambito dell'informatica umanistica, il consorzio TEI (Text Encoding Initiative, <https://tei-c.org>) ha messo a punto, a partire dal linguaggio XML, un vasto e complesso schema di codifica, proponendo una DTD (Document Type Definiton) per tutti i fenomeni dei testi umanistici. La sua finalità «è di definire uno standard di codifica, specificamente orientato alla gestione dei dati umanistico-letterari, e realizzare una normalizzazione dei formati di memorizzazione dell'informazione testuale, al fine di consentire l'interscambio dei documenti» (Francesca Tomasi, *Metodologie informatiche e discipline umanistiche*, Roma: Carocci, 2008, p. 132).

[10] Oltre a xml e ALTO è possibile esportare le immagini in formato jpeg e png o come file METS (Metadata NEcoding and Trasmission Standards) o creare documenti pdf, docx e xlsx.

[11] ALTO è uno standard sviluppato per la descrizione dei testi OCR e delle informazioni di *layout* delle pagine digitalizzate.

[12] https://gitlab.com/readcoop/transkribus/TranskribusCore/-/blob/master/src/main/resources/xsd/pagecontent_extension.xsd.

[13] In Transkribus, CER sta per Character Error Rate, ovvero tasso di errore sui caratteri. Si tratta di una metrica fondamentale per valutare l'accuratezza di un modello HTR (Handwritten Text Recognition). Cfr. <https://help.transkribus.org/it/tasso-di-errore-del-carattere-e-curva-di-apprendimento>.

[14] *Ligorio 0.3 Pyl*, in *Public AI Models in Transkribus*, <https://readcoop.eu/model/ligorio-0-3/>.

[15] *Italian Administrative Hands, 1550-1700*, in *Public AI Models in Transkribus*, <https://readcoop.eu/model/italian-administrative-hands-1550-1700/>.

MODUS

model for homogenous material, e. g. texts written by the same person or from a narrow historical period»[16].
Il modello si è mostrato inefficace nel trascrivere i testi anche nei casi in cui le immagini sottoposte a text recognition presentavano poche righe di scrittura e un'interlinea ben distanziata, come nel seguente esempio:

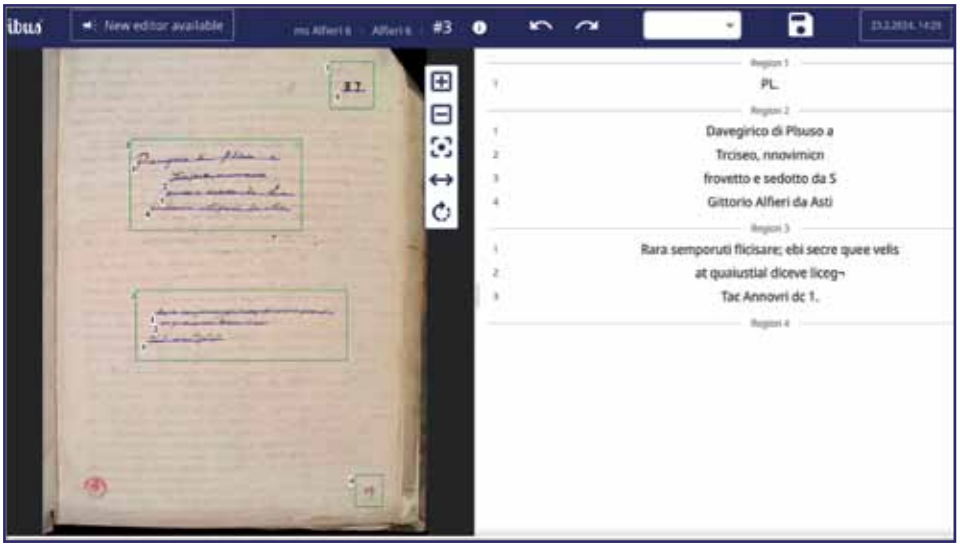

Figura 1. *Esempio di trascrizione di Alfieri 6 con modello Transkribus Italian Handwriting M1 della c. 29r. Firenze, Biblioteca Medicea Laurenziana, Ms. Alfieri 6, c. 29r. Su concessione del MiC. È vietata ogni ulteriore riproduzione con qualsiasi mezzo*

Le cause del mancato successo nel processo di trascrizione automatica possono probabilmente essere attribuite a una serie di fattori legati alla grafia dell'autore. Pur non essendo particolarmente difficile da leggere per un occhio umano, la scrittura di Alfieri presenta una serie di variabilità formali che la rendono meno adatta a una lettura automatica standardizzata. Non si tratta infatti di una grafia "standard" nel senso tecnico richiesto dai modelli HTR[17], ovvero caratterizzata da tratti ripetitivi, regolarità formale e coerenza morfologica tra le lettere. Al contrario, Alfieri mostra una scrittura soggetta a oscillazioni grafiche, abbreviazioni idiosincratiche e modifiche successive sul testo (come cancellature, sovrascritture, integrazioni a margine), tutti elementi che introducono rumore visivo e rendono più complesso l'addestramento del modello. Inoltre, la stessa disposizione del testo all'interno

[16] *Transkribus Italian Handwriting M1*, in *Public AI Models in Transkribus*, <https://readcoop.eu/model/italian-general-model/>.

[17] La "grafia standard" in Transkribus è una forma normalizzata di scrittura che fornisce il modello di riferimento per la trascrizione dei testi. Non esiste una sola "grafia standard": «There does not exist a general model for all the handwritings [...]. When choosing a text model, you need to consider the following: 1) the type of material, handwritten or printed; 2) the language; 2) the period; 3) the type of script; 4) the Character Error Rate (CER)» (Cfr. <https://help.transkribus.org/choosing-a-model>).

del manoscritto – talvolta disomogenea, con variazioni nella spaziatura e allineamenti irregolari – può compromettere l'efficacia del *layout analysis*, fase preliminare fondamentale per una corretta segmentazione delle righe e delle parole.
Si è pertanto deciso di seguire le indicazioni fornite da Transkribus, addestrando un modello privato, composto da materiale omogeneo, e stabilendo *Transkribus Italian Handwriting M1* come modello di base. Sebbene una delle *key-features* di Transkribus sia quella di permettere la creazione di modelli personalizzati e privati, generarne uno senza una base di riferimento è un'operazione molto complessa e poco significativa se non si ha un elevato numero di carte su cui addestrare il sistema: i modelli pubblici esistenti possono quindi essere utilizzati come punti di partenza per ridurre la quantità di nuovi dati necessari.
Il *corpus* di riferimento individuato è quello contenente i testi del *Panegirico di Plinio a Trajano* (cc. 29*r*-38*r*) e della *Virtù Sconosciuta* (cc. 41*r*-50*r*). Per avere un adeguato numero di carte e di parole su cui addestrare il modello, infatti, il solo materiale del *Panegirico* non è quantitativamente sufficiente[18]: si è resa pertanto necessaria l'aggiunta di materiale per ottenere un numero valido di dati su cui addestrare il modello. La scelta è ricaduta sulle carte della *Virtù Sconosciuta* per diverse ragioni. Anzitutto, le stesure del *Panegirico* e della *Virtù* contenute in Alfieri 6 risalgono a periodi contigui[19] e presentano delle grafie del tutto simili, sebbene il dialogo sia caratterizzato da un numero maggiore di correzioni e di interventi dell'autore; le opere presentano inoltre analoghe strutture di impaginazione. La relativa omogeneità nella grafia e nell'impaginazione delle pagine rende il materiale adeguato alla produzione di un modello di grafia dell'autore. Del materiale scelto, una parte delle carte del *Panegirico* è stata utilizzata per costruire la *ground truth*[20] e la restante è stata sottoposta alla *text recognition* mediante il modello personalizzato; le carte della *Virtù Sconosciuta*, invece, hanno fornito la restante parte della *ground truth* per allenare il modello.
Sono quindi state selezionate le carte per formare la *ground truth* per un totale di 13 sulle 45 disponibili[21]. L'analisi del layout è stata eseguita automaticamente, prendendo come modello *Universal Lines* e modificando eventuali imprecisioni su regioni di testo e *baseline*. I risultati dell'analisi automatica del layout si sono rivelati soddisfacenti, e il numero di interventi manuali nella segmentazione del testo è stato minimo. In questo caso, all'effi-

[18] Le linee guida Transkribus raccomandando almeno 10.000 parole per addestrare il modello di un manoscritto attribuito ad una sola mano di scrittura. «In the case of handwritten documents, our advice is to train the model on at least 10,000 words for each hand» (Cfr. *Transkribus Help Center*, <https://help.transkribus.org/data-preparation>).

[19] Stando alle annotazioni poste da Alfieri nel manoscritto, la stesura del *Panegirico* risale al marzo 1785, quella della *Virtù Sconosciuta* al gennaio 1786. Nel *Rendimento di conti* l'autore annota: «1785. Nel Marzo, ideato, e scritto d'un fiato, il *Panegirico*»; «1786. Nel Gennaio […] steso il dialogo della *Virtù Sconosciuta*» (V. Alfieri, *Rendimento di conti da darsi al Tribunal d'Apollo*, in *Opere*, v. I, cit., p. 434).

[20] La *ground truth* è la trascrizione manuale corretta di un testo manoscritto, usata come base di confronto per addestrare e valutare i modelli HTR. L'insieme di dati della *ground truth* è cruciale, in quanto i modelli di machine learning operano attraverso la replicazione statistica dei pattern forniti in fase di addestramento. Di conseguenza, la qualità e l'affidabilità della *ground truth* influenzano in maniera determinante le prestazioni e l'accuratezza del modello risultante. Cfr. anche <https://blog.transkribus.org/en/what-is-ground-truth>.

[21] Nello specifico, le carte selezionate dal *Panegirico* per costituire la *ground truth* sono: cc. 30*r*-30*v*; 31*r*-31*v*; 32*r*; 33*v*; 35*r*; 36*v*. Per la *Virtù Sconosciuta*, le carte selezionate sono: cc. 44*v*; 45*r*-*v*; 46*r*; 47*r*.

cienza di Transkribus si somma il fatto che il manoscritto oggetto di studio è una con tutta probabilità una copia in pulito della prima stesura, non presentando segni di incertezza nella struttura e nella composizione, così come nella scrittura, minuta, regolare e densa[22].
I testi delle carte scelte per la *ground truth* sono stati quindi sottoposti ad una *text recognition* mediante il modello pubblico *Transkribus Italian Handwriting M1*; le trascrizioni ottenute sono state poi corrette manualmente. La *ground truth* «is crucial for the training of probabilistic HTR models since it is the basis of their learning, that is, the training sample. For this reason, GT [*ground truth*] must be as precise as possible to obtain useful results»[23].
Si è poi proceduto all'addestramento del modello: per il *training set* sono state selezionate le 13 pagine del documento contenenti la *ground truth*, il *validation set* è invece formato dal 10% del set delle carte di training. Per il riconoscimento automatico della scrittura manoscritta è stato scelto il motore PyLaia, noto per la sua accuratezza nei compiti di Handwritten Text Recognition. Come *base model* è stato utilizzato *Transkribus Italian Handwriting M1*, già addestrato su una varietà di scritture italiane. A partire da questo, è stato avviato il processo di generazione di un modello personalizzato. In condizioni di carico server regolari, l'addestramento di un *training set* di 13 pagine con PyLaia ha richiesto circa due ore. Al termine dell'elaborazione, il sistema ha inviato una notifica di completamento tramite email.
Terminato l'addestramento, il software ha prodotto un modello con CER del 13.40%. Nello specifico, il modello presenta un CER del 13% sul *training set* e di 13.8% sul *validation set*.

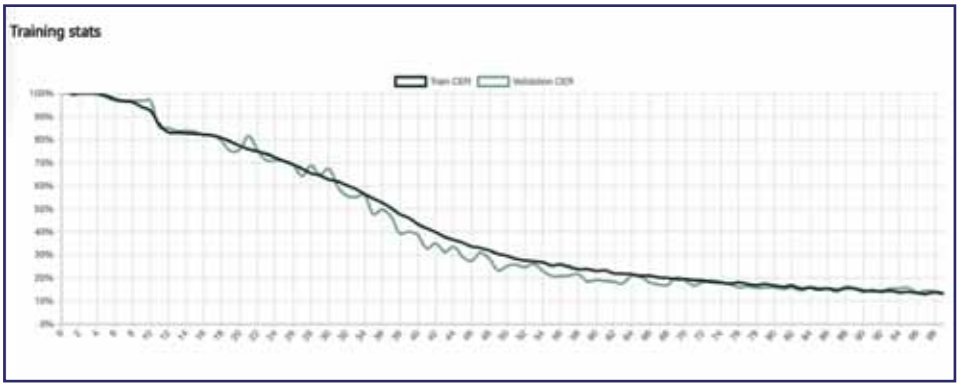

Figura 2. *Learning curve del modello* Alfieri

È stata quindi generata la trascrizione automatica di una delle carte (c. 32*v*, Fig. 4) con il modello appena creato, il risultato è poi stato confrontato con quello precedentemente ottenuto dal modello pubblico *Italian Handwriting M1*. Il modello addestrato sulle carte si è mostrato molto più performante, nonostante il CER più elevato: si propone di se-

[22] V. Alfieri, *Panegirico di Plinio a Trajano. Parigi sbastigliato. Le Mosche e l'Api*, cit., p. 18.
[23] Pepita Raventós Pajares — Celio Hernández Tornero — Simon Martin Meritxell, *AI and archive. Handwritten Text Recognition Applied to Patrimonial Holdings: An Example of 10 diaries written by Spanish Republican Teachers in 1932*, «IEEE Computer Society» (2022), p. 2572-2577: 2573-2574.

guito la stessa sezione di testo trascritta prima con il modello pubblico, poi con quello addestrato, infine il testo corretto.

Text recognition della c. 32*v* con il modello *Italian Handwriting M1*:
9 Dai meviti postri ragioni dicootanera per cui indorro il debbi a vessistno liberrà, av
10 con lieve cn pare posga e porratintina sa, e da baste quali sola; il desidevarla
11 sersomente csi dogtij e que dboto a meritarla questo parvesi sere E nsa credere gia
12 o nel su libemi altra intrasache dobbesir mopre a Frasaro, noe alle logg si
13 c poli gi diserore; ma che cessando egli porsono nella proveri daltos porenso al

Text recognition della c. 32*v* con il modello *Alfieri* (Fig. 2):
9 Dai meriti nostri cagioni dimostrarti per cui indurre ti debbi a restitaione libertà, non è
10 cosi lieve; ma pure prima, e potentisima sia, e da bastar quasi sola; il desiderarla
11 andantemente noi tutti; e gran titolo a meritarla questo parer ti dere. E non credere già,
12 che de nnel dir libertà altro intenda, che d'obbedir sempre a Trajano; sioè alle leggi, di
13 sui eggli, è difensore; ma, che, assando egli, possono nella pertena daloro potente al

Testo corretto manualmente (in grassetto le parole corrette rispetto al modello *Alfieri*):
9 Dei meriti nostri cagioni dimostrarti per cui indurre ti debbi a **restituire** libertà, non è
10 **così** lieve; ma pure prima, e **potentissima** sia, e da bastar quasi sola; il desiderarla
11 **ardentemente** noi tutti; e gran titolo a meritarla questo parer ti **deve**. E non credere già,
12 che io **nel** dir libertà altro intenda, che d'obbedir sempre a Trajano; **cioè** alle leggi, di
13 cui **egli** è difensore; ma, che, **cessando** egli, possono nella **persona d'altro** potente al

Pur rilevando un significativo miglioramento rispetto alla trascrizione con il modello pubblico M1, si è deciso di addestrare un nuovo modello aumentando il numero di fogli di *ground truth* da 13 a 24: il maggior numero di dati a disposizione per l'addestramento dovrebbe infatti fornire esiti più soddisfacenti rispetto a quelli fino ad ora ottenuti.
Definita la nuova *ground truth*, è stato allenato il modello con il motore PyLaia. Per il *training set* sono state selezionate le 24 pagine del documento contenenti le *ground truth*, il *validation set* è stato creato automaticamente sul 10% delle carte del training. È stato poi avviato il lavoro e, terminato l'allenamento, il software ha prodotto un modello con CER del 6.9%; in particolare il *train* CER è del 7.4%, il *validation* CER è del 6.9%:

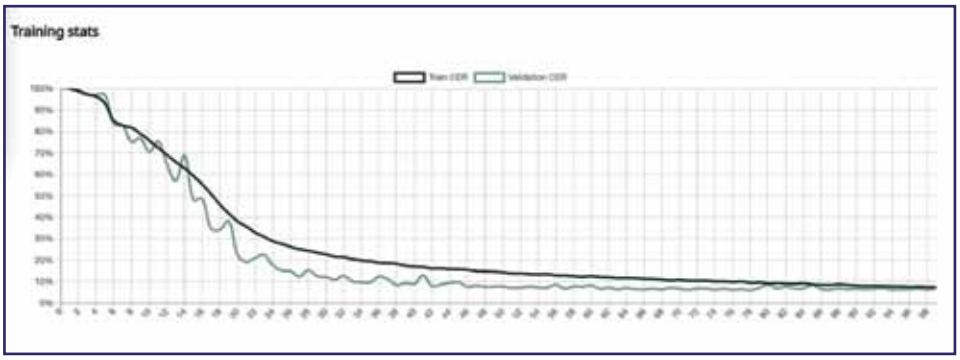

Figura 3. *Learning curve del modello Alfieri 2*

Conclusioni

Stando a quanto disponibile nelle linee guida di Transkribus, «results with a CER of 10% or below can be considered very efficient for automated transcription»[24]. Il valore del set di convalida, inoltre, è il più significativo tra i due, poiché mostra come il modello si comporta nelle pagine su cui non è stato addestrato.
Si propone di seguito la trascrizione con il modello *Alfieri 2* (Fig. 3) della sezione di testo della c. 32*v*. precedentemente analizzata:

9 Dai meriti nostri cagioni dimostrarti per cui indurre ti debbi a restituiue libertà, non e
10 così lieve; ma prure prima, e potentissima sia, e da bestar quasi sola; il desiderarla
11 ordentemente noi tutti; e gran titolo a meritarla questo parer ti deve. e non credere già
12 che de nel dir libertà altro intenda, che d'obbedir sempre a Trajano; cioè alle leggi, di
13 cue eueli è difersore; ma, che, cessando egli, possono nella persena d'aloro potente al

Lo studio condotto sul manoscritto Alfieri 6 conferma l'efficacia di Transkribus per la trascrizione automatica di testi manoscritti, evidenziando al contempo le potenzialità e i limiti delle tecnologie HTR applicate a materiali storici.
L'aumento della *ground truth* e l'addestramento di un secondo modello su un numero di dati più elevato[25] ha prodotto una trascrizione migliore rispetto alla prima: l'incremento da 13 a 24 pagine di training ha prodotto un miglioramento sostanziale delle prestazioni del modello, confermando le raccomandazioni delle linee guida Transkribus. Questo aspetto assume particolare rilevanza per i progetti di digitalizzazione di *corpus* letterari,

[24] Cfr. *Character error rate and learning curve*, <https://help.transkribus.org/character-error-rate-and-learning-curve>.
[25] La *ground truth* del modello *Alfieri* ammonta a 13 pagine con 6.071 parole, quella del modello *Alfieri 2* comprende 24 pagine e 12.148 parole.

suggerendo la necessità di pianificare investimenti adeguati nella fase di preparazione dei dati di addestramento.

Tuttavia persistono alcuni problemi, come la confusione sistematica di alcune lettere: è il caso della *q* e della *g*, della *a* e della *e*, o della *a* e della *o*, che l'autore trascrive in maniera quasi identica. Alla linea # 10, ad esempio, la *a* della parola "bastar" viene identificata erroneamente come *e;* alla linea # 11 si verifica uno scambio della *a* con la *o*, per cui la trascrizione riporta la parola "ordentemente" al posto di "ardentemente".

Per superare questi limiti, è possibile integrare le trascrizioni HTR con strategie di post-correzione automatizzata: la trascrizione prodotta dal modello verrebbe confrontata con dizionari storici o *corpora* dell'italiano settecentesco per identificare parole improbabili o errori tipici; successivamente, algoritmi di correzione (basati su regole, *fuzzy matching* o modelli linguistici statistici) suggerirebbero alternative coerenti, che potrebbero essere approvate o modificate dall'operatore umano; in questo modo si migliorerebbe la qualità delle trascrizioni, riducendo gli errori e senza richiedere correzioni manuali parola per parola.

Inoltre, la creazione di un modello specifico per la mano di Alfieri rappresenta un contributo metodologico che va oltre il singolo caso studio: il modello sviluppato costituisce infatti una risorsa riutilizzabile per future ricerche sul *corpus* alfieriano, il quale potrebbe essere implementato e aggiornato con l'aggiunta di nuovo materiale dell'autore, realizzando uno strumento sempre più efficiente e condivisibile con altri studiosi, favorendo al contempo dinamiche collaborative.

L'approccio adottato potrà infine servire da modello per progetti simili su altri autori, contribuendo alla creazione di un patrimonio digitale di strumenti specializzati per la trascrizione automatica di testi letterari storici.

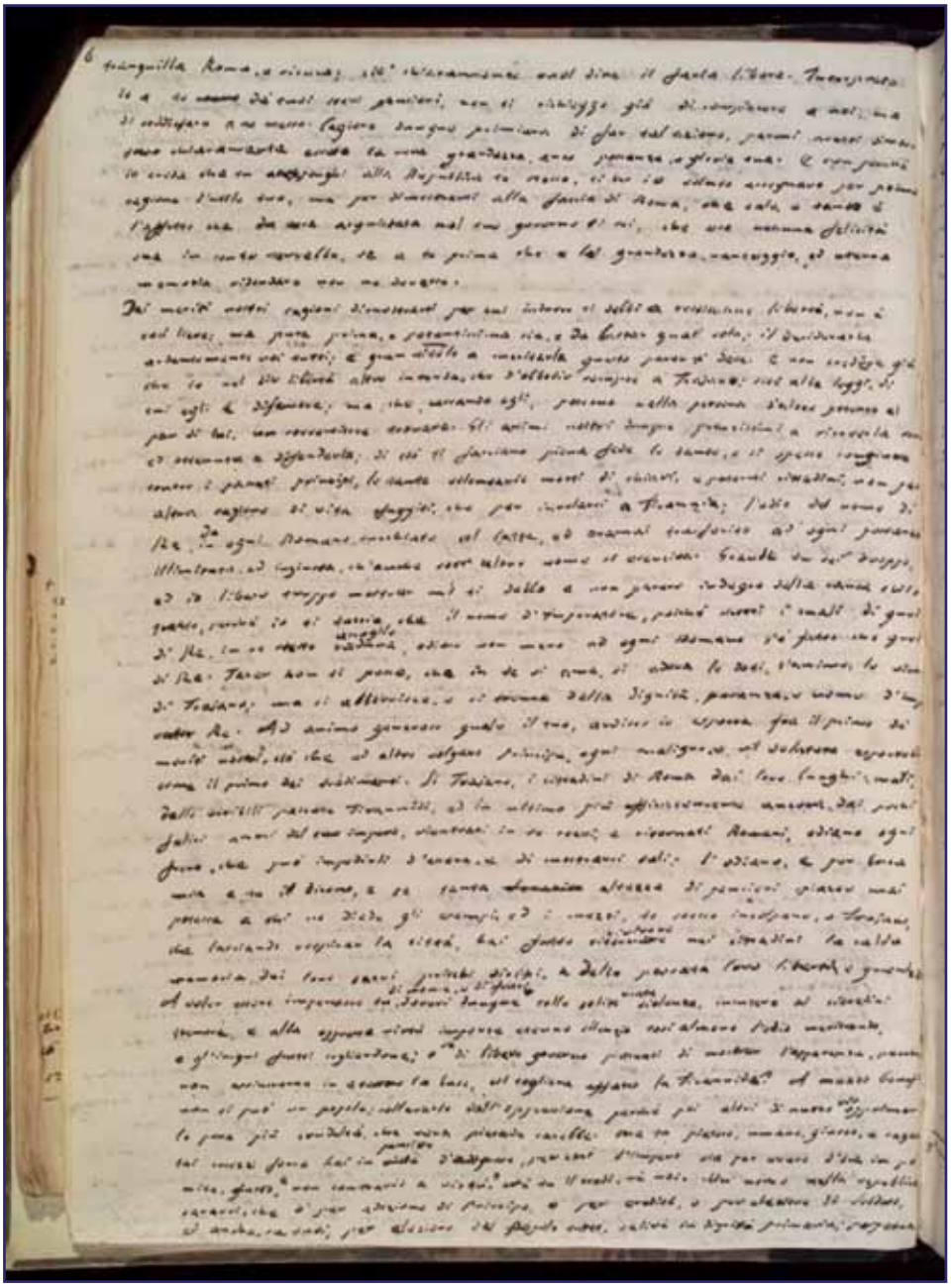

Figura 4. *Firenze, Biblioteca Medicea Laurenziana, ms. Alfieri 6, c. 32v. Su concessione del MiC. È vietata ogni ulteriore riproduzione con qualsiasi mezzo*

The study explores the use of the Transkribus platform for the automatic transcription of manuscripts, using texts included in the Alfieri 6 manuscript (Florence, Biblioteca Medicea Laurenziana). After an initial test phase with the generic Italian Handwriting M1 model, the results of which proved unsatisfactory, a customised model was created from the M1 model, which obtained a CER of 13.40%. A second training, expanding the ground truth set, reduced the CER to 6.9%, with more accurate transcriptions. Despite the progress, some difficulties in distinguishing between similar letters remain, but the study confirms the effectiveness of Transkribus for transcribing handwritten texts, suggesting the importance of adequate data to improve the quality of automatic transcription.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

Visualising Interpretive Layers in Cultural Heritage: Data narratives with Linked Open Data in mythLOD collection

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00148

Valentina Pasqual — Francesca Tomasi[1]
Digital Humanities Advanced Research Centre (/DH.arc), Department of Classical Philology and Italian Studies, University of Bologna

This work examines the evolving methods of disseminating Cultural Heritage data, emphasising the transition to Linked Open Data (LOD) and storytelling for enhanced user engagement. The mythLOD project showcases this transition by converting the Mythologiae data collection into LOD, enabling richer data interpretation and visualisation. mythLOD integrates the artwork descriptions and their scholarly interpretations, connecting artworks with literary sources through a layered data model. The resulting Knowledge Base is visualised via an online platform, making the information accessible through an online catalogue and a data storytelling interface. These visualisations reveal latent knowledge and foster deeper exploration of Cultural Heritage data, particularly by revealing the connections between literary sources and artworks and demonstrating LOD's potential to enrich user interaction and knowledge acquisition.

1. Introduction

Cultural heritage (CH) data are constantly growing from digitisation projects by many GLAM institutions. The Digital Humanities (DH) community widely adopted Linked Open Data (LOD) as a standard for knowledge organisation in the CH domain, promoting interoperability and exchange[2]. Several CH platforms disseminate their data through browsable online catalogues with faceted search interfaces, e.g. Europeana[3]. Graphic User Interfaces (GUI) make data accessible also to non-IT experts, such as museum

[1] This contribution is the product of collaboration between the authors on the mythLOD project. For the writing of this article, Francesca Tomasi is credited with writing the *Introduction* and Valentina Pasqual with writing the *Data Modelling* and *Browsing interface and data narratives* sections. The *Conclusions* section is shared.
The mythLOD project was conceived and developed within the context of the Digital Humanities Advanced Research Centre (/DH.arc, <https://centri.unibo.it/dharc/en>) research laboratory, active at the Department of Classical and Italian Philology (FICLIT, <https://ficlit.unibo.it/it>) of the University of Bologna. Special thanks go to the FrameLAB research group, creator and provider of the Mythologiae collection.
This manuscript is an extended version of a previously published contribution. Valentina Pasqual — Francesca Tomasi, *Data narratives with Linked Open Data, the case of mythLOD storytelling*. In: *Digital Humanities 2023. Collaboration as Opportunity (DH2023)*, Graz, Austria (2023). DOI: 10.5281/zenodo.8107673.

[2] Hyvönen 2012; Alexiev 2018; Po 2020.

[3] https://www.europeana.eu/it.

MODUS

curators, humanities scholars, students, or the general public, by hiding sometimes complex SPARQL queries. More recently, dissemination practices have relied on narration as a new expedient to interactively involve users more efficiently and effectively[4]. Data storytelling practices can visualise latent relations and new perspectives over data, augmenting users' knowledge. While faceted search allows users to customise data exploration depending on their individual interests, storytelling guides users through a defined narrative path. Visual analytics represents «the science of analytical reasoning facilitated by interactive visual interfaces»[5], employing visualisation and interaction methods to incorporate expert human judgment into the data analysis process. Indeed, visual analytics serves as a methodology that blends human intuition with mathematical deduction to directly perceive patterns and extract insights from complex data structures[6]. Drucker[7] argues for a reevaluation of graphical expressions in digital environments, highlighting the bias introduced by relying solely on conventional statistical data interpretation, criticising the belief that data is inherently quantitative, and emphasising the importance of recognising its qualitative, interpretive nature (referred to as *capta*). In this context, data modelling choices (e.g., ontology selection) highly influence the expressivity of data visualisation and the design of such narrative paths.

mythLOD collection[8] is the revalorisation of Mythologiae data collection[9] by converting its data from tabular data to LOD. The objective of this work is to overtake the modelling, and therefore the visualisation of its data, of mere factual data by bringing out and enhancing the interpretation of the scholars studying the Mythologiae collection and to reveal, through data visualisation expedients, hidden correlations in data. The project workflow followed the life cycle of curation for digital objects, as described by Tomasi[10], addressing the data dissemination task through two visualisations: an online catalogue browsable by facets and a data storytelling experiment applied to the Aeneid by Virgil case study.

2. Data modelling

mythLOD data have been modelled using the Digital Hermeneutics layered approach[11], which implies Nanopublication[12] structure to encode information about cultural objects and contextual information. This approach consists of four layers, and each layer is represented as a named graph in RDF format, organising specific sets of metadata[13].

4 Eide et al. 2019; Windhager et al. 2019.
5 Thomas – Cook 2006.
6 Cui 2019.
7 Drucker 2011.
8 Pasqual – Tomasi 2021.
9 https://mythologiae.unibo.it/.
10 Tomasi 2022.
11 Daquino et al. 2020.
12 Groth et al. 2010.
13 A complete overview of mythLOD data model can be found at: <https://dharc-org.github.io/mythlod/static/datamodel.html>.

- *Layer 0, factual data*. It holds cultural objects' and literary sources' descriptive metadata, forming the foundation of knowledge upon which interpretations (represented in Layer 1, assertion) are built. For instance, it includes details such as authorship, creation dates, types, periods and collocations. The FRBRoo ontology[14] is used to represent cultural objects and literary sources. Additionally, the HuCit ontology[15] is used to model the structure of classical sources, enabling a more granular representation of their content as canonical citations.

- *Layer 1, assertion*. It stores scholars' interpretations regarding cultural objects by representing the relationships between artworks, the conceptual categories they depict, the associated literary sources, and canonical citations[16]. This layer encapsulates the scholarly discourse about the collection's artworks, providing an explicit formalisation of their meaning and interrelations within the broader cultural context.

- *Layer 2, provenance*. It holds contextual information about the interpretations (represented in Layer 1, assertion), detailing their type, who made them, when, and with what criteria, mainly reusing HiCo ontology[17].

- *Layer 3, publication information*. It stores details about the publication of each Nanopublication, including who generated it and when by reusing PROV-o ontology[18].

The semantic model establishes explicit relationships between main entities - cultural objects, literary sources, scholarly interpretations, and contextual metadata - through formal ontological structures that integrate Digital Hermeneutics concepts. These relationships allow for complex queries that can traverse from factual descriptions (Layer 0) through scholarly assertions (Layer 1) to their provenance and publication information (Layers 2-3), creating a comprehensive Knowledge Base (KB) that enhances data quality and supports the identification, citation, and evaluation of scholarly arguments. Data from the Mythologiae collection has been converted into LOD format based on the aforementioned data model. Additionally, controlled forms of agents and literary sources have been used. Canonical citations have been aligned with the Perseus Digital Library's URN CTS[19]. Dates have been converted into a machine-readable format, and geospatial coordinates of the locations have been retrieved from Wikidata to visualise the instances on the map.
mythLOD KB stores 4.260 heterogeneous museum artefacts. These pieces are cat-

[14] IFLA Working Group 2017.
[15] Romanello – Pasin 2013.
[16] In other words, the assertion layer preserves information according to this pattern: "The artwork X represents the mythological scene Y, mentioned in citation Z, extracted from literary source W".
[17] Daquino – Tomasi 2015.
[18] https://www.w3.org/TR/2013/REC-prov-o-20130430/.
[19] https://sites.tufts.edu/perseusupdates/2021/01/05/what-is-a-cts-urn/.

egorised into six types (minor arts, drawings, mosaics, paintings, vase paintings, and sculptures) and span a wide range of historical and artistic contexts, featuring contributions from 1.291 different artists, currently held by 419 cultural institutions, and made across 11 distinct epochs. The artworks in mythLOD depict 132 mythological scenes, which have been organised according to an internal taxonomy. Examples include scenes such as "Enea fugge da Troia in fiamme" ("Aeneas flees from burning Troy"). Each scene is linked to one or more of 84 literary sources categorised as classical sources, medieval or modern sources, modern rewritings and screenplay rewritings, and written by 72 different authors. mythLOD thus serves as a rich repository for exploring the intersection of art, mythology, and literature across various historical periods and artistic media.

3. Browsing interface and data narratives

mythLOD data are accessible via two main visualisations, namely the catalogue and the storytelling sections on the website[20]. While offering a complete overview of the dataset, they have been involved in an actual testing activity for the whole dataset with the domain experts. The visualisations graphically represent the project competency questions, hiding the technical details (e.g., SPARQL queries). The two visualisations have been exposed to domain experts involved in Mythologiae to review both the representativeness of the mythLOD data model and the correctness of data in the KB.

3.1 Browsing interpretations through the catalogue

mythLOD catalogue[21] (Fig. 1), presents a complete overview of the artworks, which can be browsed through facet filters as their metadata, related references and scholars' interpretations. In particular, the facets contain some selected information from the first three layers[22] of the data model: cultural objects are then browsable via a set of factual metadata (e.g. collocation, period, keywords, type and author), scholars' interpretations (assertions) addressing the cultural object's representation (e.g. category) and the literary sources associated with the artworks (e.g. the reference itself, its type and author), and, finally, the interpretations' contextual metadata (e.g. interpretation type, criterion and author). Facets are organised to express the layered approach adopted by the mythLOD data model. This design aims to provide users with a comprehensive browsing interface that highlights scholars' perspectives (referred to as *capta*[23]) on the Mythologiae data, while hiding technicalities and complexities (e.g., the FRBRoo event-centric approach).

[20] https://dharc-org.github.io/mythlod/.
[21] https://dharc-org.github.io/mythlod/catalogue/.
[22] The fourth layer (Publication Information) have not been included in the visualisation since we deemed this information are more related to data exchange and interoperability with respect to users' needs.
[23] Drucker 2011.

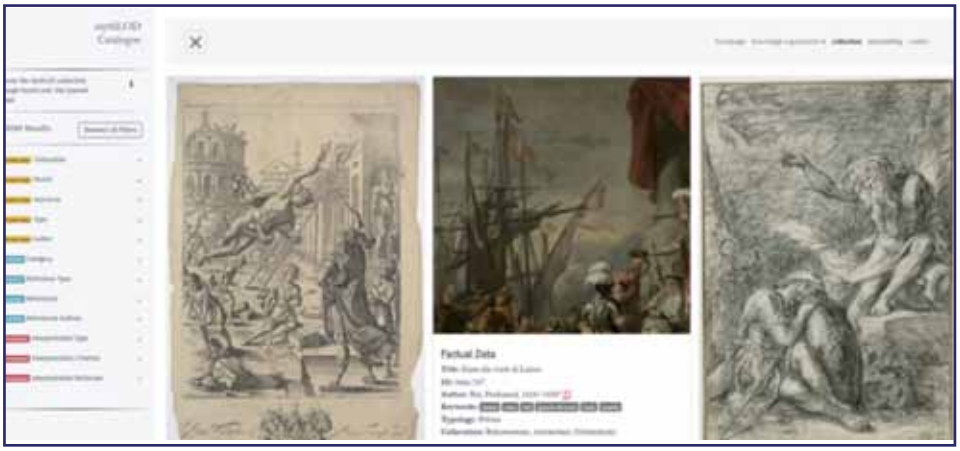

Figure 1. *The mythLOD catalogue and faceted search*

The catalogue's faceted search allows users to customise their search by adding filters from the three data layers. This browsing system allows for complex query searches on mythLOD data, for instance: all modern art (period) paintings (type) that have been interpreted to represent a mythological scene referenced by Virgil (reference author). Each artwork in the catalogue is shown with a set of metadata extracted from the first three layers of the data model. Specifically, at the assertion level (first layer), the catalogue displays a list of literary sources connected to the described artwork, supplemented with a direct link to VIAF[24] and the Perseus Digital Library (only for classical sources). This integration allows users to access the full texts for further exploration. A more accurate granularity is reserved for canonical citations, which are connected to Perseus Digital Library with the use of URN CTS to link to the specific passage describing the mythological scene depicted by the artwork. Overall, the processes of LOD curation (and in particular, data modelling choices, data cleaning, entity reconciliation and metadata enrichment) show its implications at the visualisation level, allowing users to perform more complex queries concerning Mythologiae catalogue[25] and creating interactive relations between mythLOD data and external sources as Perseus Digital Library to deepen the data exploration.

3.2 Telling Aeneis through museal art

While the catalogue section primarily focuses on artworks, *Telling Aeneis through museal art* is the data storytelling section about Aeneid contents in the mythLOD collection. The narrative consists of a restricted set of visualisations investigating the relation between mythLOD artworks and Aeneid literary reference by re-proposing the project competency questions into four categories.

[24] https://sites.tufts.edu/perseusupdates/2021/01/05/what-is-a-cts-urn/.
[25] https://mythologiae.unibo.it/index.php/ricerca/.

- Artworks' temporality (*when*): a timeline illustrates when the museum works concerning the themes of the Aeneid were produced

- Artworks' spatiality (*where*): a map shows where the museum artworks concerning the themes of the Aeneid are located

- Artworks' contents (*what*): a heatmap represents the most frequently cited Aeneid passages in the artworks depicted scenes; a word cloud represents the most frequent terms implied in the description of the artworks related to Aeneid; a network represents the literary references which co-occur with Aeneid in the dataset (Fig. 2).

- Agents involved with artworks (*who*): a network shows the relations between authors who created an artwork whose depicted scene has also been recognised to be present in the Aeneid.

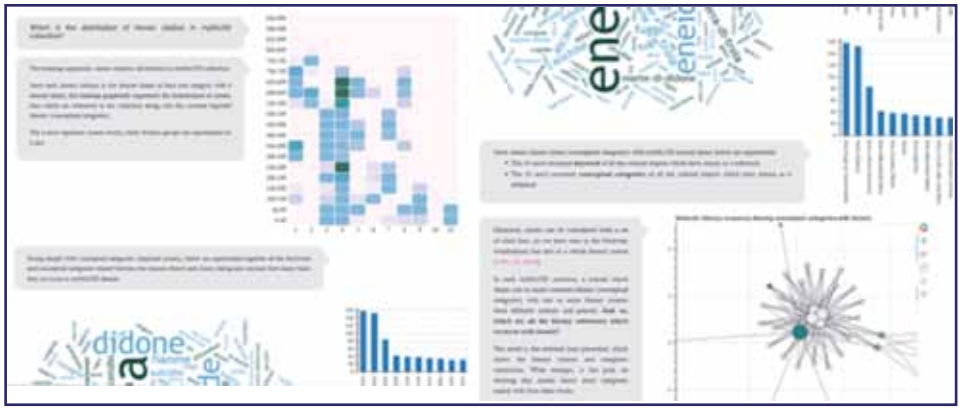

Figure 2. *Snapshots of contents visualisation (what) in mythLOD storytelling*

In addition to the four categories of interest, a textual introduction and tips for each visualisation have been added to guide the user through the four sections. In particular, each tip reports the competency question to which the visualisation is answering and a brief explanation of the results. For instance, the network visualisation in the "WHAT" section addresses the question: "Which literary references co-occur with Aeneid?". The visualisation is also accompanied by a brief explanation noting that four literary sources (namely, Francesco Petrarch's "Trionfi", Giuseppe Ungaretti's "Vita di un Uomo", Dante Alighieri's "Divina Commedia", and Giacomo Leopardi's "Canti") exhibit the highest degree of co-occurrence.

Visualisations strictly depend on the entire data curation process, such as knowledge organisation (e.g., adopted data models and ontologies), data cleaning, and alignment with external resources. For example, the granular representation of canonical citations

is presented in the *heatmap* (Fig. 2), highlighting Aeneid's most cited lines and motives in mythLOD KB. Additionally, machine-readable dates were necessary to create the timeline. At the same time, geospatial coordinates extracted from Wikidata were used to visualise the instances on the map, fostering the valorisation of spatio-temporal data.

4. Conclusions

The implementation of the mythLOD project demonstrates the transformative potential of LOD in CH knowledge representation and production. The analysis presented in this work reveals an epistemological shift facilitated by the transition from traditional tabular data structures (*Mythologiae* data) to a semantically enriched LOD framework (mythLOD KB). This transformation repositions the narrative focus from artifact-centric cataloguing to source-oriented knowledge networks, inverting the traditional museum paradigm.

The visualisation interfaces — both the faceted browsing catalogue and the Aeneid-focused narrative pathway — exemplify how Semantic Web technologies enable the manifestation of latent semantic relationships within cultural datasets. The four-layered Digital Hermeneutics approach, particularly through its integration of factual metadata with scholarly interpretation via the Nanopublication framework, constitutes a methodological advancement beyond simple digitization toward interpretive augmentation of cultural heritage digital resources.

Quantitave analysis of the mythLOD KB — comprehending 4.260 artifacts, 132 mythological scenes, and 84 literary sources — reveals that visualisation methodologies serve not merely as passive interfaces but as active analytical instruments that reify complex relationships between museal artworks and literary references. The Aeneid case study demonstrates how visualisation can function as a hermeneutic tool, allowing for the identification of patterns in artistic interpretation across temporal and spatial dimensions that would remain latent in traditional representations.

Furthermore, this research contributes to ongoing scholarly discourse on what Drucker[26] terms *capta* — the recognition that cultural data embodies interpretive judgments rather than mere objective facts. The mythLOD platform operationalises this theoretical position by explicitly modelling and visualising scholarly interpretations as first-class knowledge entities, thereby acknowledging the subjective dimension of CH information.

These findings suggest that LOD implementations in CH domains should be conceptualised not merely as technical infrastructure but as epistemological frameworks that fundamentally alter how knowledge is organised, accessed, and interpreted. The semantic enrichment facilitated by LOD creates conditions for what might be termed as "emergent knowledge" — insights that exist only as potential within traditional data structures but become accessible through semantically informed visualization interfaces. This work ultimately positions LOD visualisation not as a supplementary feature of digital CH but as a transformative methodology that reconfigures the relationship between artefacts, sources, and interpretation.

[26] Drucker 2011.

MODUS

Questo lavoro analizza i metodi per la diffusione dei dati del patrimonio culturale, enfatizzando la transizione verso i Linked Open Data (LOD) e lo storytelling per migliorare il coinvolgimento degli utenti. Il progetto mythLOD illustra questa transizione convertendo la collezione di dati Mythologiae in LOD, permettendo un'interpretazione e una visualizzazione dei dati più ricche. mythLOD integra le descrizioni delle opere d'arte con le loro interpretazioni, collegando le opere alle fonti letterarie attraverso un modello dati a livelli. La base di conoscenza risultante è visualizzata tramite una piattaforma online, che rende le informazioni accessibili attraverso un catalogo digitale e un'interfaccia di narrazione dei dati. Queste visualizzazioni rivelano conoscenze latenti e favoriscono un'esplorazione più profonda dei dati del patrimonio culturale, in particolare mettendo in luce le connessioni tra fonti letterarie e opere d'arte, dimostrando il potenziale dei LOD nell'arricchire l'interazione e l'acquisizione di conoscenza da parte degli utenti.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

References

Alexiev 2018 — Vladimir Alexiev. *Museum Linked Open Data: Ontologies, Datasets, Projects*. «Digital Presentation and Preservation of Cultural and Scientific Heritage», 8 (2018), p. 19–50. <https://doi.org/10.3233/ISU-2010-0613>.

Daquino et al. 2020 — Marilena Daquino — Valentina Pasqual — Francesca Tomasi. *Knowledge Representation of Digital Hermeneutics of Archival and Literary Sources*. «JLIS.It», 11 (2020), n. 3, p. 59–76. <https://doi.org/10.4403/jlis.it-12642>.

Daquino – Tomasi 2015 — Marilena Daquino — Francesca Tomasi. *Historical Context Ontology (HiCO): a Conceptual Model for Describing Context Information of Cultural Heritage Objects.* In: *Metadata and Semantics Research: 9th research conference, MTSR 2015, Manchester, UK, September 9-11*, 2015. (Communications in Computer and Information Science; 544). Cham: Springer International Publishing, 2015. <https://doi.org/10.1007/978-3-319-24129-6_37>.

Eide et al. 2019 — Øyvind Eide — Zoe Schubert — Enes Türkoğlu — Jan Gerrit Wieners — Kai Niebes. *The intangibility of tangible objects: re-telling artefact stories through spatial multimedia annotations and 3D objects.* In: *ICOM Kyoto 2019, 25th ICOM General Conference, 1-7 September 2019*. Kyoto, Japan, 2019. <https://doi.org/10.5281/zenodo.3878966>.

Groth et al. 2010 — Paul Groth — Andrew Gibson — Jan Velterop. *The anatomy of a nanopublication. «Information Services & Use»*, 30 (2010), n. 1-2, p. 51-56. <https://doi.org/10.3233/ISU-2010-0613>.

Hyvönen 2012 — Eero Hyvönen. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web.* In: *Synthesis Lectures on Data, Semantics, and Knowledge (SLDSK).* Cham: Springer International Publishing, 2012. <https://doi.org/10.1007/978-3-031-79438-4>.

IFLA Working Group 2017 — IFLA Working Group on FRBR/CRM Dialogue. *Definition of FRBRoo: a conceptual model for bibliographic information in object-oriented formalism*. International Federation of Library Associations and Institutions (IFLA), 2017. <https://repository.ifla.org/handle/20.500.14598/659>.

Pasqual — Tomasi 2022 — Valentina Pasqual — Francesca Tomasi. *Linked open data per la valorizzazione di collezioni culturali: il dataset mythLOD*. «AIB studi», 62 (2022), n. 1, p. 149–68. <https://doi.org/10.2426/aibstudi-13301>.

Po et al. 2022 — Laura Po — Nikos Bikakis — Federico Desimoni — George Papastefanatos. *Linked data visualization: techniques, tools, and big data*. Morgan & Claypool Publishers, 2020.

Romanello — Pasin 2013 — Matteo Romanello — Michele Pasin. *Citations and Annotations in Classics: Old Problems and New Perspectives*. In: *DH-CASE '13. Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: Metadata, Vocabularies and Techniques in the Digital Humanities*. New York: Association for Computing Machinery, 2013, p. 1–8. <https://doi.org/10.1145/2517978.2517981>.

Thomas — Cook 2006 — Jim Joshua Thomas — Kris A. Cook. *A visual analytics agenda*. «IEEE Computer Graphics and Applications», 26 (2006), n. 1, p. 10-13. <https://doi.org/10.1109/MCG.2006.5>.

Tomasi 2022 — Francesca Tomasi. *Organizzare la conoscenza: digital humanities e web semantico*. Milano: Editrice Bibliografica, 2022

Windhager et al. 2019 — Florian Windhager — Paolo Federico — Günther Schreder — Katrin Glinka — Marian Dörk — Silvia Miksch. *Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges*. «IEEE Transactions on Visualization and Computer Graphics», 25 (2019), n. 6, p. 2311-2330. <https://doi.org/10.1109/TVCG.2018.2830759>.

Cultural Heritage Accessibility through Multisensory Interactions - Updates from the MuseIT project

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00149

Nasrine Olson — *The Swedish School of Library and Information Science, University of Borås*
Flavia Massara — *The Central Institute for the Union Catalogue of Italian Libraries (ICCU)*
Vasilis Papadopoulos — *Catalink Limited (CTL)*
Maria Kyrou — *Information Technologies Institute, Centre for Research & Technology Hellas (CERTH)*
Albert Meroño-Peñuela — *King's College London (KCL)*
Moa Johansson — *ShareMusic & Performing Arts*
Vyacheslav Tykhonov — *Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Sciences, (DANS-KNAW)*
James Hanlon — *X-System Ltd*[1]

As the MuseIT project (Multisensory, User-centred, Shared Cultural Experiences through Interactive Technologies) reaches its final stage, this article offers an overview of its key achievements in promoting accessibility and inclusion in cultural heritage. Funded by the Horizon Europe programme, MuseIT brought together a diverse consortium of researchers, technologists, and cultural stakeholders to develop innovative tools for inclusive digital culture. By leveraging multimodal interfaces, immersive technologies, and AI-based accessibility solutions, the project enabled more meaningful cultural experiences for individuals with sensory, cognitive, and physical disabilities. This contribution highlights the main outcomes of MuseIT, emphasizing its interdisciplinary methods, community involvement, and ethical considerations, while laying the groundwork for future progress in inclusive technology and cultural engagement; a particular focus is dedicated to the event organized within the framework of the project, held in Rome on May 14, 2025, at the National Central Library of Rome, titled "Heritage Without Barriers".

Introduction

Multisensory, User-centred, Shared Cultural Experiences through Interactive Technologies (MuseIT)[2] is a three-year project (2022-2025) funded under Horizon Europe, coordinated by the University of Borås, Sweden. The consortium comprises around 70 members across 12 partner organisations in nine countries. The project aims to advance inclusive and multisensory access to cultural heritage through participatory design, multimodal technologies, immersive experiences, and promotion of improved policies.

[1] Other authors who contributed to the paper are Thomas Van Erven, Sándor Darányi, Konstantinos Avgerinakis, Georgia Georgiou, Eleftherios Anastasovitis, Spiros Nikolopoulos, Sophia Alexandersson, Nigel Osborne, Kim Ferguson, Nitisha Jain, Andrea Scharnhorst.

[2] https://www.muse-it.eu/.

MODUS

An earlier overview of the project's initial activities and perspectives was published in DigItalia in 2023[3].

As the MuseIT project nears its conclusion, by this report, we reflect on a journey defined by innovate on, inclusion, and interdisciplinary collaboration. Funded under the Horizon Europe framework, MuseIT set out with an ambitious goal: to make cultural heritage more accessible and inclusive through the use of advanced technologies, particularly for and with individuals with sensory, cognitive, and physical disabilities.

Since its inception, MuseIT has brought together researchers, technologists, artists, and stakeholders from across Europe to reimagine how people experience and engage with cultural content. Through a combination of cutting-edge solutions –including multimodal interfaces, immersive experiences, and AI-driven accessibility tools– the project has developed a comprehensive approach to inclusive digital culture.

This report outlines the main achievements of MuseIT across its core areas. From novel interaction paradigms and sensory translation systems to community engagement strategies and ethical frameworks, each section reflects our commitment to broadening access and participation in the cultural sphere. As the project draws to a close, we present these results not only as outcomes, but as foundations for future initiatives in inclusive technology and digital heritage.

Understanding user needs – a Participatory and Inclusive Approach

At the core of MuseIT's methodology is a firm commitment to participatory design and inclusive research. From its outset, the project recognised that the meaningful inclusion of people with disabilities in the development process is essential –not only to ensure accessibility but to empower users as co-creators of cultural experiences. To this end, MuseIT has adopted a participatory co-design approach, working closely with potential users to shape technologies and experiences that genuinely meet their needs.

This approach has been grounded in established frameworks such as Universal Design and Inclusive Design and has actively involved users at different stages of the design and development process. Through over 60 participatory engagements –including workshops, interviews, symposia, and collaborative fieldwork– the project has strived to form an informed understanding of user needs across diverse groups and contexts.

MuseIT's engagements were guided by values of equality, empowerment, and democratic collaboration. Participants contributed insights into the barriers they face when engaging with cultural content, and helped co-design solutions using multisensory and digital technologies, including virtual reality (VR) and AI-driven interfaces. These interactions have not only informed technical development but also enhanced awareness within the consortium of the lived experiences of the user communities.

[3] Maud Ntonga — Juliette Pokorny — Nasrine Olson, *The MuseIT Project: co-designing inclusive technologies for better access to culture*, «DigItalia. Rivista del digitale nei beni culturali», 18 (2023), n. 1, p. 187-190, <https://digitalia.cultura.gov.it/article/view/3009>.

MODUS

In particular, co-design sessions allowed users to directly shape design concepts, test early prototypes, and propose refinements. This collaborative process has helped ensure that the outcomes are not only technically effective but socially meaningful and emotionally resonant. The participatory methodology also reinforced the project's commitment to ethical research practices and the responsible development of inclusive technologies.

A key component of this process involved an interview study with individuals from the project's target user groups, which illuminated both challenges and opportunities in engaging with cultural content. Participants described difficulties accessing cultural experiences –both physical and digital– due to poorly designed environments, lack of sensory accommodations, or overly complex navigation structures. Many emphasised how cultural spaces often default to "normalised" experiences that exclude or overlook diverse modes of perception and interaction. At the same time, participants expressed a strong desire to connect with culture in ways that reflect their abilities, interests, and preferences. Opportunities identified included the potential of multisensory media, such as touch, sound, and movement-based interactions, to enable more immersive and meaningful experiences. Others noted the empowering role of co-creation itself: not only as a means of shaping technology, but as a form of cultural participation in its own right. These insights highlighted once again the importance of designing with –not just for– users, and reinforced MuseIT's mission to foster inclusive cultural futures through collaborative innovation.

While MuseIT has made significant progress in developing inclusive digital solutions, it is important to acknowledge that several of the technologies produced remain at the prototype stage. Not all tools are currently fully accessible, and further refinement is needed to achieve universal usability. However, the project has established a solid conceptual and technical foundation; building blocks that can be improved, integrated with other systems, or scaled in future efforts. These prototypes[4] serve as proof of concept, offering valuable insights and demonstrating clear pathways for how participatory design and inclusive technologies can evolve together to support broader accessibility goals. Next to the prototypes, MuseIT has embraced standard solutions for research data management and long-term digital archiving, resulting in various Dataverse-based repository instances which will be hosted by consortium members also after the duration of MuseIT.

Technical Innovations for Inclusion

Building on the insights gained from participatory work, MuseIT has developed several innovative tools that translate inclusive design into technological solutions. Below, we outline some of these developments.

[4] For a better understanding of the developed prototypes, the training videos available at the following link can be consulted: <https://www.youtube.com/watch?v=BPN835oq5O4&list=PL11mJw4sA9UPMVYQ8CK_LzyP3CscJSCbQ>.

MODUS

HaptiVerse: Enabling Meaningful Communication Through Touch

One of the technologies developed in the project is a system of interrelated components that enable design of and communication of meaning-bearing haptic messages. The HaptiVerse system[5] represents a novel approach to haptic communication, developed with a focus on inclusion, accessibility, and co-design for use with and by the deafblind community. Building on existing practices such as Tactile Sign Language and Social-Haptic Communication (SHC), HaptiVerse introduces a modular framework for creating, transmitting, and receiving meaning-bearing haptic signals. These signals can be used to convey structured information through touch –something especially vital for individuals with dual sensory impairments. At its core, HaptiVerse consists of several interconnected components: HaptiDesigner (a tool for designing haptic patterns or "haptograms"), HaptiMux (a routing and communication hub), HaptiBoard (hardware for actuator control), HaptiMesh/HaptiWear (wearable interfaces), and HaptioTek (a shared, expandable library of haptic signs). This integrated ecosystem supports real-time, remote, and one-to-many communication, enabling individuals to receive haptic messages even without direct physical contact with a human sender. The system's flexibility is a defining feature. It accommodates different user needs– whether they wish to design custom patterns, integrate their own devices, or simply stream haptic content. HaptiVerse is designed not only for assistive communication but also for experimentation and co-creation in broader contexts, such as cultural settings, education, and research.

Importantly, the development of HaptiVerse has been grounded in participatory design, informed by lived experiences of people with deafblindness. Co-design sessions and field testing helped identify key challenges, such as the need for low-latency communication, modular hardware, and intuitive pattern design. These insights have directly shaped the system architecture and usage scenarios. While HaptiVerse is still in an experimental phase, its modularity and open design make it a promising foundation for future work. As the field of haptic communication matures and standards evolve, the system is built to adapt. It also lays the groundwork for a growing, community-driven haptic vocabulary-a step towards what may eventually become a tactile language.

Music Generation Based on the User's Experience in the Frontend

The goal is to create a personalized musical piece that reflects both the user's mood and the cultural heritage assets they explored on the MuseIT platform.

As the user navigates through the MuseIT interface, they can explore various cultural heritage assets. There is an option to activate their camera by clicking a button, which starts capturing their facial expressions to detect mood. The user can stop the recording whenever they choose.

[5] The HaptiVerse system is a software solution allowing for digital design of haptic devices, haptograms, and construction of basic sentences, aimed to iterative development with Social Haptic Communication.

Once the recording stops, the system processes the video using a machine learning model that analyses facial expressions and predicts the user's mood over time. At the same time, the system keeps track of the artefacts the user visited. These are sent to the MuseIT Knowledge Graph, which provides meaningful insights about each artefact. This combined data –user mood and visited artefacts– is then sent to a music generation tool that creates a custom musical piece representing the user's overall experience. The generated music is influenced by two main factors. First, the emotional tone of the music is shaped by the user's mood. For example, if the model detects mostly happy expressions, the music follows a major key with uplifting characteristics. For sadder moods, the music shifts to a minor key, producing a more melancholic feel.

Second, the cultural context of the artefacts influences the musical style. For example, if the user explored a modern art artefact, the music might include electrical instruments, or if the artefacts are from the ancient period, the composition may use instruments like the lyre or aulos. By blending these two aspects-emotion and cultural context-the result is a unique, personalized musical piece.

This technology has been tested in the first pilot of the MuseIT project in Rome. The overall feedback was really promising: people enjoyed the experience and found it engaging. Many users mentioned that the music captured their overall experience exceptionally well, and they expressed a strong interest in trying it again. However, there is still room for improvement to further enhance its usability and impact.

Haptification & sonification of artworks

The sonification and haptification track for MuseIT is based around the translation of artwork to physical sensation. By using computer vision models to extract concepts from the paintings, using nearest-neighbour algorithms to simplify colors to a predetermined scheme and mapping them with color, further by haptifying concepts and contours, we aim to give participants with limited sight an impression of what is in the painting.

For inspiration for color conversion, we looked at Scriabin's color circle, in which the composer associates musical notes with specific colors. For instance; the color red is associated with the note C, with lighter shades of red responding to a higher note. Traditionally, 12 notes are associated with shades of color –by identifying 5 shades per color, we distinguish 60 unique colors and associated notes. By converting each pixel in the artwork to a shade of these colors, and loading the result into the Picture 2 Notes application[6], a user can explore the color structure of the painting through hearing rather than vision.

A haptic mode is also supported; we express the current pixel hovered over by means of an RGBA frequency, which manifests on three actuators activating on a connected

[6] The Picture 2 Notes application is a simple JavaScript application, which can produce musical notes and haptic feedback when a user interacts with an image loaded into it.

haptic device. This allows one to sense by vibration whether a color is red, or green, or blue, with the intensity of the vibration corresponding to the proximity of the color to pure red, green, and blue.

Conceptualization of the contents of a painting was done by a computer vision model, which describes the contents of the paintings in detailed text. This data was stored and publicly published in the MuseIT Dataverse[7], which makes this effort FAIR-compliant[8].

After generating the data, Haptiverse's backend was used to stream the data to a connected vest. By converting the detailed text to Haptic Subject Index[9], an innovative approach to use simple language to describe features on paintings, it is possible to transmit the haptic patterns (which take the form of a series of four clock faces, where each of the eight respective compass directions in a clock face correspond to a single content feature), conceptually describing what is in the painting. Finally, after the contours of shapes in a selected painting were extracted programmatically, we colored them in a single pure color (which is rare in paintings). This approach allows one to express them haptically as well. By drawing the cursor closer or further away from shape contours, the respective haptic intensity of the device is increased vs. decreased, allowing one to sense this kind of information by vibration.

Multimodal Transformations with Generative tools

The CUBE-MT[10] dataset and benchmark was designed as part of the project to support more inclusive access to cultural heritage using generative AI. It includes artefacts from three cultural domains –cuisine, art, and landmarks– across eight culturally diverse countries: Brazil, France, India, Italy, Japan, Nigeria, Turkey, and the USA. For each artefact, we created six different formats –images, text, speech, music, Braille, and 3D models– to meet a range of sensory and cognitive needs, especially for people with disabilities.

The content was generated using multimodal AI tools from Hugging Face, including models like Stable Diffusion[11] (images), FastSpeech 2[12] (speech), MusicGen[13] (music), and Hunyuan3D[14] (3D models). All these models can generate multimodal assets based on prompts and other forms of input: for example, Stable Diffusion can create images from textual prompts, whereas Hunyuan3D can generate printable 3D representations

[7] MuseIT used Dataverse and created various instances, later in the footnote 13 around the ShareMusic case a reference is given to the Dataverse project as such, feel free to cross-reference; also, a good reference is Mercè Crosas, *Cloud Dataverse: A Data Repository Platform for the* Cloud, «CIO Review Open Stack», 2017, <https://openstack.cioreview.com/cxoinsight/cloud-dataverse-a-data-repository-platform-for-the-cloud-nid-24199-cid-120.html>.

[8] FAIR: Findable, Accessible, Interoperable, Reusable: <https://www.go-fair.org/fair-principles/>.

[9] Haptic Subject Index, or HSI, is a simplified schema of 4 x 8 wind-rose directions to communicate specific subject information about the artwork.

[10] https://github.com/albertmeronyo/CUBE-MT.

[11] https://huggingface.co/stabilityai/stable-diffusion-3-medium-diffusers.

[12] https://huggingface.co/facebook/fastspeech2-en-ljspeech.

[13] https://huggingface.co/facebook/musicgen-small.

[14] https://huggingface.co/tencent/Hunyuan3D-2.

using images as input. This setup allowed us to produce diverse and culturally grounded representations efficiently. Each modality serves a different purpose. Images and 3D models offer helpful visual and tactile cues for people with language or literacy challenges. Speech and Braille support users with visual impairments, while music can convey cultural context when other forms are less accessible.
To evaluate the dataset, we ran a workshop with people with aphasia, a condition that affects language. Participants engaged with the artefacts one modality at a time. Images and 3D models were particularly effective, helping users recognise and interpret content without needing to read or listen. The feedback highlighted the value of multimodal content in improving accessibility. The full dataset is openly available and integrated into the MuseIT platform for further research, design, and inclusive cultural applications.

Figure 1. *Image from a workshop focused on people with aphasia for evaluating the usefulness of multimodal representations*

Virtual museum and affective computing

MuseIT Virtual Museum is a fully immersive and multisensory eXtended Reality (XR) experience exhibiting cultural assets. This digital collection of artefacts was developed in collaboration with the Central Institute for the Union Catalogue of Italian Libraries (ICCU) of the Italian Ministry of Culture. Within the simulated and multi-thematic gallery, faithfully mirroring the scale, layout, and visual detail of a physical exhibition, visitors don Virtual Reality equipment and navigate freely among digital 2D and 3D representations of the exhibits. The experience is enhanced with audio and experimental haptic interactions, which not only enrich the user's museum experience but mainly, they facilitate accessibility. Various approaches for enhanced accessibility to the content

of the virtual museum are applied, based on the Centre for Research and Technology Hellas (CERTH) proposed framework[15].

We explored how VR technology can enable remote access to cultural experiences by creating high-quality virtual environments that closely resemble the physical world. Advanced 3D graphics render these spaces with high realism, while multisensory features (audio and haptic) reinforce inclusivity. Concurrently, an affective-computing protocol captures participants' emotional responses throughout the self-paced VR session. Each visitor wears a non-invasive electroencephalogram (EEG) device to monitor brain activity and a wrist-worn sensor to record heart rate and skin conductance, ensuring that physiological data align precisely with every interaction in the virtual space. After the session, recorded signals can be analysed to reveal how individuals emotionally engage with digital heritage environments. This methodology applies neuroscience and biomedical sensing to cultural heritage, moving beyond traditional surveys to uncover subtle, non-conscious reactions. The resulting qualitative assessment can inform adjustments to haptic intensity, audio pacing, and spatial design, thereby enhancing emotional resonance and universal accessibility in future iterations of the MuseIT Virtual Museum.

Figure 2. *A participant experiencing the Virtual Museum workstation during the Pilot 1 event demonstration in Rome*

[15] Eleftherios Anastasovitis — Georgia Georgiou — Eleni Matinopoulou — Spiros Nikolopoulos — Ioannis Kompatsiaris — Manous Roumeliotis, *Enhanced Inclusion through Advanced Immersion in Cultural Heritage: A Holistic Framework in Virtual Museology*, «Electronics», 13 (2024), n. 7, 1396. <https://doi.org/10.3390/electronics13071396>.

MODUS

A Knowledge Base for Arts and Inclusion

As demonstrated above digital technology is increasingly able to capture aspects of multimodality. But, to document multimodality aspects of cultural practices and products for the long-term remains a challenge. To create a repository which both contains artistic products from the performing arts in their multimodal nature but also supports the access to those digitized traces was another aspiration of MuseIT. For the collection of the ShareMusic[16] which entails a wide range of *objects* from scientific articles to recorded performances, a repository solution (based on the open-source Harvard Dataverse[17] archival software) was created, equipped with specific ontologies and integrated into the webspace of information around ShareMusic.

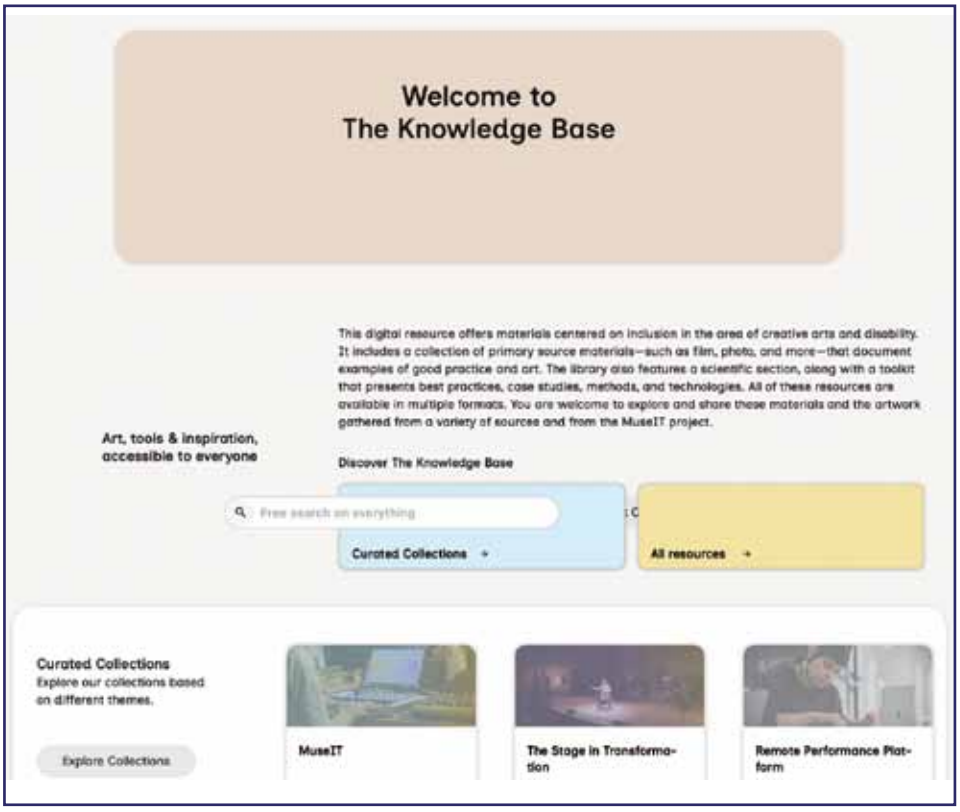

Figure 3. *Snapshot of the new user interface to the Knowledge Base for Art and Inclusion, <https://knowledgebase.sharemusic.se/>*

[16] Share Music & Performing Art is a project partner of MuseIT: <https://www.sharemusic.se/en/home>.

[17] The Dataverse Project, <https://dataverse.org/> is an open-source research data repository software - developed by the Institute for Quantitative Social Science at the Harvard University and supported and used by many other institutions. Vyacheslav Tykhonov is one of the "ambassadors" of this open software community, and at DANS as part of the <SSHOC.eu> project, a workflow to deploy Dataverse was developed. Marion Wittenberg — Vyacheslav Tykhonov — Eko Indarto — Wilko Steinhoff — Laura Huis in 't Veld — Stefan Kasberger — Philipp Conzett — Cesare Concordia — Peter Kiraly — Tomasz Parkoła, *D5.5 'Archive in a Box' repository software and proof of concept of centralised installation in the cloud*, «Zenodo», 2022, <https://doi.org/10.5281/zenodo.6676391>.

AI based tools and workflows have been tested and implemented for the co-design of the Knowledge Organisation System[18] at the backend, but also for a web interface which meets various user needs. By choosing an Open Science based Information engineering pathway, the forward compatibility of the current solution is guaranteed while at the same time, standards in long-term preservation such as FAIR principles are adhered to[19,20].

Engagement Activities

MuseIT has adopted a holistic approach, recognising that technological development alone is not sufficient to address the complex challenges of accessibility and inclusion. Instead, the project combines innovation with broader societal engagement. In addition to developing advanced technologies, MuseIT has actively contributed to policy studies, formulated actionable recommendations, and facilitated roundtable discussions. The project also prioritises awareness-raising, community engagement, and widespread dissemination of information to ensure that inclusive practices are understood, adopted, and sustained across sectors.

Building on this approach, the project has actively incorporated user engagement through co-creation and co-design methodologies and organised a wide range of engagement activities to involve users, experts, and stakeholders at multiple levels. These stakeholders include cultural institutions, disability organisations, researchers, technology providers, policy bodies, and user communities with whom the consortium had already established long-term relationships prior to the project. The network has since been broadened through targeted outreach, public events, and communications, ensuring diverse perspectives are represented. Our holistic approach and extensive user engagement efforts have not only supported the co-creation of inclusive technologies but also helped bridge the gap between research, policy, and real-world practice.

MuseIT has carried out more than 60 user engagement activities, spanning a wide spectrum of formats and levels of participation. From awareness-raising events and public symposia to in-depth co-design workshops, policy dialogues, and collaborative fieldwork, the project has actively involved diverse user groups, experts, and stakeholders. By drawing on established models, MuseIT has ensured that users are not merely recipients of information, but valued contributors and co-creators. These activities have

[18] Knowledge Organisation System is a technical term to describe any structural model to organise knowledge. Examples are thesauri, classification systems, and ontologies. For a definition see <https://www.isko.org/cyclo/>. In the case of the ShareMusic Knowledge Base, the KOS combines standard bibliographic information with specific facets describing the multimodality of some content and ways to access content.

[19] Moa Johansson — Vyacheslav Tykhonov — Sophia Alexandersson — Kim Ferguson — James Hanlon — Andrea Scharnhorst — Nigel Osborne. *A Knowledge Base for Arts and Inclusion: The Dataverse Data Archival Platform as a Knowledge Base Management System Enabling Multimodal Accessibility,* in: *Human-Centered Design, Operation and Evaluation of Mobile Communications. HCII 2025*, ed. By June Wei, George Margetis, (Lecture Notes in Computer Science; 15824), Cham: Springer, 2025, p. 291-309, <https://doi.org/10.1007/978-3-031-93064-5_19>; preprint: <https://arxiv.org/abs/2504.05976>.

[20] Dataverse at the backend: <https://database.sharemusic.se/>.

MODUS

also helped build internal capacity and a cross-disciplinary understanding within the consortium through hands-on learning and interdisciplinary dialogue. An outline of one such engagement activity is provided in the next section.

As the project approaches its conclusion, several significant events remain. These include the final MuseIT symposium, "Beyond Boundaries — Multisensory Innovation for Inclusive Futures" which took place in two parts on 2 September 2025 at the University of Borås and on 3 September at Röhsska Museum in Gothenburg. Following that event, the project team also took part in the Digital Heritage International Congress 2025[21], and hosted the second ACM Europe Summer School on Accessible and Inclusive Technologies[22], which took place in Borås on 16–19 June 2025. These final activities have subsequently extended the project's reach, deepened collaboration, and highlighted its contributions to inclusive, human-centred innovation.

Heritage without barriers: Pilot1 event demonstration in Rome

One of the most engaging recent moments of public outreach was the event held in Rome on May 14, "Heritage Without Barriers. The MuseIT project for an Inclusive Cultural Experience through Technology". The initiative was organized within the framework of the MuseIT project in collaboration with the Italian Ministry of Culture (MiC) – Central Institute for the Union Catalogue of Italian Libraries (ICCU) and the National Central Library of Rome.

The day began with institutional greetings from Paola Passarelli, General Director for Libraries and Cultural Institutes. This was followed by remarks from Stefano Campagnolo, Director of the National Central Library of Rome, and Giuliano Genetasio, Director of the Central Institute for the Union Catalogue of Italian Libraries, ICCU. Nasrine Olson, associate professor at the University of Borås and coordinator of the MuseIT project, then presented the project, highlighting its main challenges and key outcomes.

Next, Daniela Bottegoni and Alessia Varricchio from the Museo Tattile Statale Omero shared insights into their institution –an exemplary case of multisensory engagement with cultural heritage. Finally, Claretta Caroppo and Lorenzo Barello illustrated the practices of accessible performances and other inclusive initiatives promoted by the Teatro Stabile di Torino.

A roundtable discussion followed, moderated by Corinne Szteinsznaider from the Michael Culture Association. The panel brought together Cristina Da Milano (Centro Europeo per l'Organizzazione e il Management Culturale, ECCOM), Maddalena Battaggia and Lucia Sardo (Study Group on Inclusion of the Italian Library Association), Gabriella Cetorelli (UNESCO Office – Italian Ministry of Culture), and Dino Angelaccio (National Observatory on the Condition of Persons with Disabilities).

The discussion focused on the policy briefs published by the MuseIT project and subsequently expanded to address European Union policies on the topic, those implemented by the Italian Ministry of Culture and, more broadly, at the national level, as well as

[21] https://digitalheritage2025.unisi.it/.

[22] https://europe.acm.org/seasonal-schools/accessible-inclusive-technologies.

specific actions undertaken by the Italian Library Association (AIB) in the field of library accessibility.

Afterwards, the testing phase of the technologies developed within the MuseIT project was launched. Multiple workstations were set up, allowing participants to explore the platform and experiment with various tools, including haptic technologies. Five dedicated workstations were available: WebData, Haptiverse, Haptics, Virtual Museum, and Music Generation. These were managed by project partners with the support of facilitators[23].

In parallel with the technology testing phase, attendees –divided into two rotating groups– took part in a workshop focused on the accessibility of cultural heritage. Participants were invited to share their experiences, both positive and negative, and to express their perspectives in a non-judgmental and attentive environment.

This contribution aims to present the main findings that emerged from the workshop.

To break the ice and help participants feel at ease, they were invited to write their name on a post-it notes, introduce themselves, and instinctively choose a word they associated with the day so far. This activity resulted in two meaningful word clouds –each reflecting the responses of one group. While both were composed of positive terms, the differences between the two highlight how the topic can evoke a range of emotions and be interpreted through multiple perspectives.

Figure 4. *Word cloud from the first workshop*

Participants were then asked to share their negative experiences regarding the accessibility of cultural heritage. One of the first topics to come up –and one of the most passionately discussed– was the use of language and terminology. When addressing the topic of disability, it is essential to use language carefully, as even seemingly neutral terms –such as "inclusion"– may reflect a perspective in which a majority seeks to include a minority. For this reason, many participants suggested that the term "equality" –as stated in Article 3 of the Italian Constitution– would be more appropriate.

A similar reflection was made on the expression "differently abled", which was criticized for shifting the focus away from the person and for being inherently discriminatory, as it implicitly raises the question: *differently abled compared to*

[23] At each workstation, after completing the tests, users were asked to answer a set of evaluation questions. The results of this feedback will be published in project deliverable D7.3. *MuseIT Pilot Demonstrators 2nd Iterations*.

Figure 5. *Word cloud from the second workshop*

whom? Words are not neutral, and participants agreed on the importance of critically considering the language used when discussing disability.

Another issue that emerged was the subtle yet pervasive discomfort with disability that can be perceived even in environments where accessibility is ostensibly considered a priority. A particularly telling example was shared by a participant, who recounted how museum staff denied a guided tour for a group of individuals with cognitive disabilities, despite being informed that the visit was part of a pilot project for designing a tailored guide. At the same time, other tours were taking place, and the group was instead asked to follow the exhibition in reverse order.

These paradoxical experiences served as a starting point for a broader discussion on the importance of co-designing services together with persons with disabilities. Unfortunately, this crucial phase –essential for addressing the actual needs of the intended users– is often only partially implemented, covering isolated segments of the process. Participants highlighted a lack of continuity, leading to paradoxical outcomes: for instance, services that require the presence of a companion, thus failing to ensure user autonomy, or architectural solutions intended to improve accessibility that instead create new barriers, such as a door handle that narrows the passageway and ultimately renders a space inaccessible. As a result, some participants suggested placing more emphasis on making artworks accessible, rather than focusing on pathways, sharing a common frustration with the current approaches.

The "exceptionality" of these processes is a sign that sensitivity to the issue is changing, but the effort made is not enough. A cultural shift is needed; we must "be born into multiplicity" and, above all, offer an experience that is *complex and fulfilling for everyone equally: an equitable experience*, as one participant emphasizes.

To achieve a change in mindset, participants emphasized the need for what one referred to as a "minority exercise" –an intentional effort to place oneself in the position of others in order to better understand how to ensure no one is left behind. Empathy and the experience of exclusion were central themes: some participants shared that their awareness significantly increased after they personally encountered temporary disability due to an accident or illness, which allowed them to experience firsthand the barriers others face daily.

A participant shared a successful professional experience in which, to address complexity, the museum she works for organized different teams for specific disabilities, involving people with disabilities and various professionals. This approach allowed for the consideration of aspects that would otherwise have been difficult to foresee.
Another issue raised by participants was related to technology, which, in line with the previous discussion, is often designed to assist people with disabilities but is left unsupported afterward with no adequate training and awareness for those who are meant to use and manage it. Technology is undoubtedly a valuable tool that has allowed, and continues to allow, both people with disabilities and others to facilitate communication and engage in reading and cultural experiences with a certain degree of autonomy. However, on its own, it is not enough.
Later, during the workshops, participants were invited to share their positive experiences regarding accessibility. After an initial phase in which it seemed almost more difficult to recall and share positive episodes (despite the words chosen at the beginning of the meeting), many stories filled with humanity and emotion emerged.
Participants shared experiences such as the feeling of "being at home" in an exemplary museum where hospitality is genuinely extended to everyone; of theatre subscriptions gladly purchased by people with disabilities when performances are truly accessible; and of sensory experiences like "readings in the dark" designed to foster empathy and perspective-taking. Notably, it was emphasized that feeling part of society often begins with small gestures, for example, something as simple yet meaningful as receiving a wedding invitation written in Braille.
A clear and powerful idea emerged: the *right to beauty [and access to culture] does not belong to just a few*. If cultural services –and, by extension, all services– are truly designed with everyone in mind, then *the richness that stems from diversity benefits everyone*. This workshop experience, in enabling participants to speak, understand, and truly listen to one another, ultimately left everyone enriched.

Concluding Remarks

As the MuseIT project approaches its conclusion, this report presents some of the collective efforts that have shaped its course. Bringing together diverse partners, disciplines, and user communities, the project has aimed to explore how technology might support more inclusive and engaging encounters with cultural heritage. The work completed so far suggests promising directions.
MuseIT has provided a space for experimenting with multisensory technologies, participatory methods, and inclusive design principles. Through co-design and dialogue with individuals with a wide range of abilities, the project has aimed to form a better understanding of user perspectives and the barriers people face, creating the opportunity for collaborative efforts to help address them.
We acknowledge that much remains to be done. Nonetheless, we hope that the approaches taken and the insights gained will prove useful to others working in the fields of accessibility, digital heritage, and inclusive technology.

In vista della conclusione del progetto MuseIT (Multisensory, User-centred, Shared Cultural Experiences through Interactive Technologies), questo articolo presenta una panoramica dei principali risultati raggiunti nell'ambito dell'accessibilità e dell'inclusione del patrimonio culturale. Finanziato dal programma Horizon Europe, MuseIT ha riunito un consorzio eterogeneo di ricercatori, tecnologi e attori culturali per sviluppare strumenti innovativi per una cultura digitale inclusiva. Attraverso l'impiego di interfacce multimodali, tecnologie immersive e soluzioni di accessibilità basate sull'intelligenza artificiale, il progetto ha reso le esperienze culturali più significative e accessibili per persone con disabilità sensoriali, cognitive e fisiche. Questo contributo illustra i principali risultati del progetto, mettendo in evidenza i suoi approcci interdisciplinari, il coinvolgimento delle comunità e le riflessioni etiche, offrendo una base solida per futuri sviluppi nel campo della tecnologia inclusiva e della fruizione culturale; un particolare focus è dedicato all'evento organizzato nell'ambito del progetto, a Roma il 14 maggio 2025 alla Biblioteca Nazionale Centrale di Roma, "Patrimonio senza barriere".

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

Documenti e discussioni

Alfabeto arabo e OCR open source: un'analisi grafica e linguistica dei risultati di elaborazione per il recupero di dati catalografici

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00150

Riccardo Amerigo Vigliermo
Università di Modena e Reggio Emilia – FSCIRE

*Il riconoscimento ottico dei caratteri (*Optical Character Recognition, OCR*) è uno strumento centrale per la conservazione di dati digitalizzati, ma mostra limiti significativi con alfabeti non latini, specie di fronte a stampe desuete o manoscritti, dove dati "rumorosi" e variabili contestuali ostacolano il riconoscimento ottimale. L'OCR è centrale nello sviluppo di applicazioni che prevedono l'impiego di altre tecniche di elaborazione del linguaggio naturale (*Natural Language Processing, NLP*). L'analisi degli errori, come fase di post-elaborazione (*post-processing*), può migliorare l'accuratezza soprattutto se combinata con un'analisi contestuale. Lo studio qui presentato ha l'obiettivo di delineare i tratti comuni degli errori commessi da tali OCR testati nel progetto Digital Maktaba.*

Introduzione

Gli OCR mostrano ancora limiti nell'elaborazione di testi non nativi digitali, specialmente in ambito storico-religioso, a causa della natura eterogenea e "rumorosa" dei dati e delle difficoltà linguistiche, in particolare se applicati all'alfabeto arabo[1]. Spesso i risultati generati dai sistemi OCR si rivelano infatti inesatti, inficiando potenzialmente la validità di applicazioni di tecniche di *Natural Language Processing* (NLP), che utilizzano come dati di partenza i risultati degli OCR per compiti anche molto diversi tra loro[2]. In questo senso, studi recenti hanno indicato aree di ricerca promettenti concentrandosi in particolare sulle fasi di preparazione dei dati (*pre-processing*) e della loro correzione e rifinitura (*post-processing*). Ricerche recenti hanno indicato nei processi di *pre-* e *post-processing* una via per migliorare l'accuratezza, tramite correzioni contestuali, modelli probabilistici, soluzioni basate su regole (*rule-based*)[3]. A queste strategie di ottimizzazione si sono recentemente aggiunte le possibilità offerte dai nuovi sviluppi nel campo dei *Large Language Models* (LLM) e *Large Multimodal Models* (LMM) - le cosiddette Intelligenze artificiali - i cui studi hanno individuato nell'integrazione fra OCR e LLM

[1] Ahmed — Abidi 2019.
[2] Chiron et al. 2017.
[3] Nguyen et al. 2021, p. 9; Bassil — Alwani 2012; Afli — Loïc —Schwenk 2016; Khosrobeygi et al. 2020.

un punto di partenza di grande potenziale per il miglioramento dei processi di recupero delle informazioni[4].

Lo studio qui presentato vuole porre il focus sull'importanza dell'analisi dei risultati in contesti testuali multilingua[5] e ha l'obiettivo di delineare i tratti comuni degli errori commessi da alcuni sistemi OCR open-source testati nel corso del progetto Digital Maktaba (da qui in avanti DM), nato dalla collaborazione tra informatici, storici, bibliotecari, ingegneri e linguisti riuniti dalla start-up mim.fscire, dall'Università di Modena e Reggio Emilia (UniMoRe) e dalla Fondazione per le scienze religiose (FSCIRE)[6].

Il set di dati alla base di questo progetto proviene dalla vasta collezione di documenti in formato PDF messa a disposizione dalla Biblioteca "Giorgio La Pira" di Palermo, specializzata in Storia e dottrine dell'Islam e facente parte della FSCIRE. Sono stati condotti diversi test con soluzioni OCR open-source, per poter stabilire quale potesse fornire i migliori risultati su un sottoinsieme di frontespizi e copertine multilingua in alfabeto arabo (una selezione di 100 esemplari tra frontespizi e copertine, con vari *layout* e con diversi stili calligrafici scansionati in formato PDF). I frontespizi sono stati trattati con l'impiego di GoogleDocs[7], Tesseract[8] e EasyOCR[9], generando i risultati di seguito illustrati[10].

Brevi note sulle caratteristiche generali della scrittura araba

Alcune caratteristiche generali della scrittura araba possono essere così riassunte:

– L'alfabeto arabo è esclusivamente corsivo con andamento da destra a sinistra[11].

– L'alfabeto è esclusivamente consonantico e omografico (si veda sotto). Le vocali (fatḥa "a", ḍamma "u" e kasra "i") e i punti diacritici sono espressi sopra o sotto il rigo. Le vocali lunghe (ā, ī, ū) sono sistematicamente rappresentate mediante i grafemi ا (alif), ي (yā', y) e و (wāw, w)[12].

– I grafemi sono dipendenti dalla posizione in cui compaiono nella parola e modificano la loro forma se sono isolati, in posizione iniziale, mediana o finale (es. ع عـ ـعـ ـع).

– I grafemi ا و ر ز د ذ anche ژ in persiano) non legano a sinistra, creando sequenze grafiche non connesse da legature (ad esempio ضرب , أضرار).

[4] Boros et al. 2024; Pakhale 2023.

[5] Smith — Cordell 2018.

[6] Da novembre 2022 fa parte del progetto ITSERR finanziato dal Ministero della ricerca italiano con i fondi del programma NextGenerationEU. L'obiettivo principale di DM è quello di sviluppare un sistema per l'estrapolazione automatica della conoscenza e la classificazione di documenti in alfabeti non latini (in particolare in arabo) in contesti di vasti archivi di documenti digitali (in formato PDF).

[7] https://christophersrose.com/2020/05/05/how-to-use-google-docs-to-ocr Arabic-text/.

[8] https://tesseract-ocr.github.io/tessdoc/.

[9] https://github.com/JaidedAI/EasyOCR/blob/master/README.md.

[10] I sistemi OCR open source hanno limitazioni di addestramento sui font, causando inesattezze con caratteri nuovi o misti (Zoizou — Zarghili — Chaker 2020, p. 576). Per questo motivo "errore" sarà utilizzato tra virgolette.

[11] Se si considera la resa delle cifre la scrittura può considerarsi anche bidirezionale, in quanto i numeri seguono un orientamento inverso. La compresenza di grafemi arabi e numeri può generare conflitti di lettura bidirezionale, aspetto assai rilevante per l'estrazione di metadati catalografici dai frontespizi. Ulteriori complicazioni possono derivare dalla coesistenza di numeri latini (سنة 1345), arabi (سنة ١٣٤٥) e persiani (سال ١٣٤٥), con set grafici parzialmente diversi e codici Unicode distinti: numeri arabi (0660-0669) e numeri persiani (06F0-06F9), nonostante non tutte le cifre siano graficamente differenti.

[12] Garbini — Durand 1994.

MODUS

– Alcune legature tipiche di alcuni stili calligrafici comportano modifiche nelle posizioni dei punti diacritici come anche del rigo, come accade ad esempio per lo stile *nastaʿliq*[13].

Di seguito si fornisce una tavola per la traslitterazione doppia basata rispettivamente sullo standard ISO 233-2:1993 per l'arabo e ISO 233-3:2023 per il persiano. Eventuali discostamenti da questi standard adottati nel presente studio sono indicati in nota.

Arabo	traslitterazione	Persiano	traslitterazione
ا	v.note	ا	v.note
ء	ʿ (v.note)	ب	b
ب	b	پ	p
ت	t	ت	t
ث	ṯ	ث	ṯ
ج	ǧ	ج	ǧ
ح	ḥ	چ	č
خ	ḫ	ح	ḥ
د	d	خ	ḫ
ذ	ḏ	د	d
ر	r	ذ	ẕ
ز	z	ر	r
س	s	ز	z
ش	š	ژ	ž
ص	ṣ	س	s
ض	ḍ	ش	š
ط	ṭ	ص	ṣ
ظ	ẓ	ض	ẓ (v.note)
ع	ʿ	ط	ṭ
غ	ġ	ظ	ẓ
ف	f	ع	ʿ
ق	q	غ	ġ
ك	k	ف	f
ل	l	ق	q
م	m	ک	k
ن	n	گ	g
ه	h	ل	l
ة	v. note	م	m
و	w/ū	ن	n
ي	y/ī	و	v
ى	v. note	ه	h
		ى	y

[13] Il riconoscimento OCR dello stile *nastaʿliq* presenta criticità specifiche dovute alla diagonalità e alle legature che si sovrappongono sia orizzontalmente che verticalmente. La segmentazione richiede quindi un'analisi bidimensionale che tenga in considerazione sovrapposizioni di caratteri adiacenti e legature (Javed — Hussain 2009, p. 2), nonché righi multipli orizzontali e diagonali, fattori che aumentando la complessità del riconoscimento.

MODUS

1. La *alif* ا se preceduta da vocale breve <a> segnala l'allungamento di quest'ultima e sarà traslitterata come <ā>.
2. L'articolo determinativo *al-*, quando presente, se preceduto dalle preposizioni *bi*, *li*, *ʿalā*, *ʿilā*, *fī* e le congiunzioni *wa* e *fa*, verrà riportato come segue: es. *wa-l-*, *fī-l-*, *bi-l-* ecc.
3. Nel caso di particelle come *bi*, *li*, *ʿalā*, *ʿilā*, *fī*, se seguite da un pronome enclitico, non saranno separate da quest'ultimo mediante un trattino (es. *fiha*, *ʿalayna*, *ʿilayhim* ecc.).
4. La *tāʾ marbūṭa* verrà traslitterata con *-a* (finale); in caso di stato costrutto (*iḍāfa*) o dopo una *alif* di prolungamento *ā* verrà traslitterata con *t* (es. *ṣalāt*).
5. La *alif madda* آ e la *alif maqṣūra* ى verranno traslitterate con *ā*.
6. Nel caso in cui il nome inizi in *alif-hamza* <ʾ>, quest'ultima sarà omessa e verrà resa maiuscola la vocale adiacente, come ad esempio in: أبو *Abū*. Nel caso in cui il nome inizi in *ʿayn* <ʿ> verrà resa maiuscola la vocale adiacente successiva.
7. Per quanto riguarda la traslitterazione del grafema ض con <ẓ> in ISO 233-3: 2023 (persiano), nel presente studio si è mantenuta la <ż> della versione precedente ISO 233-3:1999.

Similarità e omografia

Da un punto di vista terminologico è necessario precisare due concetti per una migliore comprensione delle relazioni grafiche tra i vari grafemi. Si tratta di relazioni scontate ad un occhio "umano", ma da non sottovalutare nel contesto del riconoscimento "automatico".

Il primo concetto è quello di similarità dei grafemi. In questo caso due grafemi possono condividere parti del tratto calligrafico[14], sopra o sotto al rigo, che se non perfettamente riprodotte possono portare a confondere un grafema per l'altro, (per esempio: <*d*> د e <*r*> ر). Il secondo riguarda invece l'omografia[15], caratteristica tipica dell'alfabeto arabo che vede due o più grafemi, generalmente legati in una stringa (o parola), condividere uno o più glifi. I grafemi vengono distinti solamente mediante l'utilizzo di uno, due o tre punti diacritici, come ad esempio: <*ṭ*> ط e <*ẓ*> ظ. In questo caso, i punti diacritici sono l'unico fattore discriminante tra caratteri, di conseguenza il loro mancato riconoscimento può generare forme omografe valide a livello linguistico ma non esatte dal punto di vista del risultato, come anche forme non esistenti.

[14] Per *glifo* si intende qui l'unità grafica minima, distinta dal *grafema*, che ha invece valore linguistico; la precisazione è necessaria poiché non tutti i glifi sono grafemi. Con il termine "carattere" invece si indicherà l'insieme di glifi e grafemi in un particolare stile calligrafico e tipografico (al pari del termine inglese *font*).

[15] L'omografia era già nota ai grammatici arabi con il nome di ištirāk lafẓī o ištirāk lafẓī ḫaṭṭī, si veda a tal proposito (Munaǧǧid 1999, p. 30). Gli studi distinguono generalmente l'omografia in eterografa e omofona: la prima occorre quando vengono a mancare i segni diacritici di vocalizzazione (es. كتب kataba, kattaba, kutiba ecc.), la seconda invece quando due parole identiche e con la stessa vocalizzazione recano due significati diversi (بيت bayt = "casa" o "verso di poesia") (Taouka — Coltheart 2004, p. 32).

Metodologia

Dal punto di vista operativo, i frontespizi sono stati automaticamente estratti e convertiti in formato PNG utilizzando il modulo Fitz della libreria PymuPDF[16], preservando la risoluzione originale. Tale conversione si è resa necessaria poiché i sistemi OCR open-source analizzati non supportano input PDF, ad eccezione di GoogleDocs.
La scelta dell'impiego dei frontespizi e delle copertine per formare il set di campioni è stata operata seguendo i seguenti criteri:

– presenza/assenza di frontespizio o copertina;
– presenza di sfondi di vario colore;
– varietà nella qualità delle immagini pdf;
– varietà nello stato di deterioramento dei testi;
– varietà di "rumore" presente sul testo;
– rappresentazione proporzionale delle lingue dell'archivio (arabo, persiano, turco azero)[17];
– rappresentazione della varietà di stili calligrafici arabi utilizzati in diversi volumi, in diverse pagine e anche nella medesima pagina;
– presenza di testo vocalizzato/non vocalizzato;

Dal momento che la natura dei risultati dell'elaborazione di un sistema OCR è soggetta a numerose variabili, gli esempi proposti e la loro categorizzazione hanno valore puramente dimostrativo, senza pretendere di definire esaustivamente tutti i possibili esiti. Data la fluidità della calligrafia araba, mappare ogni errore sarebbe poco produttivo; si analizzano invece gli errori più distinguibili e frequenti tramite una categorizzazione *ad hoc*. Ad ogni categoria corrisponde una tabella dove la colonna di destra rappresenterà il carattere di input (Cinp) mentre la colonna di sinistra riporta l'output (Cout). L'analisi considera due livelli di variabili: "interne" (relative a riconoscimento di singolo grafema e sistema OCR) ed "esterne" (estranee ma ugualmente influenti sul riconoscimento).

È inoltre necessario tenere in considerazione una serie di variabili grafematiche, come:

– differenze tra vari alfabeti a base arabo (ad esempio persiano);
– carattere non nativo digitale / carattere nativo digitale (es. Tasmeem Emiri)[18];
– i caratteri seguono gli stili calligrafici arabo-persiano (*nasḫ*, *nastaʿlīq* - *taʿlīq*, *kūfī*, *ruqʿa*, *diwānī*, *ṯuluṯ*, *muḥaqqaq*, *tawqīʿ*)[19];
– i caratteri "artistici" sono quelli ideati appositamente per un particolare frontespizio (ogni carattere ha peculiarità grafiche proprie che comportano numerose sfide a livello grafico);

[16] https://pymupdf.readthedocs.io/en/latest/recipes-images.html.
[17] La presente analisi si concentrerà principalmente sull'arabo in quanto alfabeto d'origine e in comune con persiano e turco azero iraniano. Sporadici esempi in persiano si troveranno nelle tabelle.
[18] Font del progetto Tasmeem costruito con le specifiche ACE (Arabic Calligraphic Engine) di DecoType. Emiri è una ricostruzione dello stile tipografico dell'edizione del Corano del Cairo del 1923 (Milo 2006).
[19] Per un'analisi complessiva degli stili calligrafici arabi vedano gli studi di al-Ǧubūrī (1998, p. 203-222) e Schimmel — Rivolta 1992 (p. 32–33).

– presenza di segni diacritici di vocalizzazione (sopra o sotto il rigo);
– presenza di decorazioni estranee al carattere che generano "rumore" confondendosi talvolta con vocalizzazione o con gli stessi punti diacritici.

Variabili legate ai sistemi OCR open source disponibili sono invece:
– qualità di riconoscimento complessiva;
– qualità di riconoscimento di metadati sintattici (ad esempio la posizione del testo nel *layout* della pagina);
– grado di addestramento del sistema sull'alfabeto arabo;
– robustezza alle variazioni di dimensione e di font.

Per quanto riguarda le variabili "esterne", le più rilevanti sono:
– qualità dell'immagine complessiva;
– risoluzione del carattere;
– variazioni del carattere in stile, grandezza e forma;
– colori dello sfondo e dei dati presenti sulla copertina (qualora non sia presente il frontespizio);
– presenza di alterazioni sulla pagina del testo (macchie, timbri ecc.)

Emergono poi ulteriori variabili "strutturali" non collegate al riconoscimento OCR ma legate a questioni più specificatamente biblioteconomiche, di struttura del libro, che risultano rilevanti in un progetto come DM, quali: struttura del frontespizio, scelte artistiche nella resa degli elementi (titolo, autore, editore) e posizione del testo nel *layout*. Si tratta di variabili che riguardano l'analisi dell'impaginazione, il riconoscimento delle porzioni testuali e dei metadati catalografici.
Le sezioni successive introducono schemi di traslitterazione, concetti di similarità e omografia, studi sulla valutazione OCR e analisi degli "errori" per categorie, inclusi esempi di scrittura coranica.

Studi sulla valutazione di sistemi OCR: caratteristiche principali e diversità con la presente analisi

La valutazione OCR può essere condotta in modalità *blackbox* o *whitebox*: la prima considera il sistema come un tutt'uno, confrontando *output* e *ground truth* (cioè fra risultato del riconoscimento e dati veri e verificati[20]), la seconda invece analizza le prestazioni dei singoli moduli. Secondo diversi studi, le metriche possono variare per livello di dettaglio: accuratezza di grafemi e parole, precisione, esattezza su stringhe e classi di grafemi[21]. Per l'arabo, sono state proposte classificazioni basate su posizione, punti diacritici, grafemi con hamza (o "hamzati" a cui si aggiunge il grafema

[20] Kanungo — Marton — Bulbul 1999.
[21] Saber — Ahmed — Hadhoud 2014; Batawi — Abulnaja 2012; Rice — Kanai — Nartker 1993.

ك <*k*>)[22], rapporto col rigo e presenza di tratti chiusi come nel caso della classe *loop*, la quale raccoglie tutti i grafemi che presentano un tratto chiuso circolare o simile (es. ط و ع ﻫ)[23].

Le metriche più comuni e generalmente diffuse sono Character Error Rate (CER) e Word Error Rate (WER): il primo misura i caratteri errati rispetto al totale, il secondo prende in considerazione le parole. Entrambi si basano sulla distanza di Levenshtein, cioè il numero minimo di operazioni base (sostituzione, cancellazione e inserimento) necessarie per correggere l'OCR, e di conseguenza allineare il testo estratto alla *ground truth*[24].

I sistemi presi in considerazione nella presente analisi sono stati valutati sia con le metriche note (CER e WER), sia con metriche *ad hoc* (qdiff, qscore) introdotte per valutare la qualità dei risultati in relazione a input e *ground truth* del set di dati preso in considerazione.

Dal momento che i parametri valutativi generalmente utilizzati per i sistemi OCR (tra cui l'accuratezza) poco si confacevano con le variabili presenti nei campioni a disposizione, sono state ideate due metriche distinte dipendenti dalla qualità del documento e basate principalmente sul riscontro linguistico:

– qdiff (range [-2,2]), relazione di qualità dei risultati rispetto alla *ground truth*;
– qscore (range [1,5]) qualità del risultato di riconoscimento rispetto a qualità di input.

Con le seguenti formule:

– qscore = $(5 - ((2 - oq) * (iq + 1))$ se $oq \neq 0{,}1$ altrimenti,
– qdiff = $oq - iq$

La metrica qdiff penalizza i risultati OCR quando la qualità di *input* supera quella di *output*. I test evidenziano che Tesseract ed EasyOCR estraggono testi di qualità media fornendo metadati di posizione, mentre Google Docs offre una qualità superiore e identificazione linguistica ma senza metadati sulla posizione del testo[25].

I test hanno evidenziato diversi punti di forza (Tesseract, EasyOCR: estrazione media e metadati di layout; Google Docs: riconoscimento linguistico e qualità più alta) e alcuni limiti (assenza di metadati di posizione). Gli esperimenti condotti tramite le metriche sopra descritte hanno fornito indicazioni sull'implementazione delle librerie OCR *off-the-shelf* (cioè pronte all'uso) testate per l'impiego nel contesto catalografico, senza però approfondire l'aspetto grafico e linguistico. Per questa ragione il presente studio si soffermerà proprio sull'analisi grafica e sulle implicazioni linguistiche derivanti dal riconoscimento del testo. L'esito del processo è infatti da considerarsi, in questo contesto, come una parte di un più ampio sforzo finalizzato allo sviluppo di uno strumento di catalogazione semi-automatica per una biblioteca digitale.

[22] Saber et al. 2016, p. 453.
[23] Nel presente studio tale tratto verrà chiamato: "occhiello".
[24] Alghamdi — Alkhazi — Teahan 2016, p. 2.
[25] Per i risultati numerici e statistici dei test condotti in DM si rimanda a Bergamaschi et al. 2022, p. 12–13.

Analisi degli errori e loro categorizzazione. Scambio di forme con alto grado di similarità

Cout per Cinp	**Cout**	**Cinp**
d per *r*	د	ر
ḏ per *z*	ذ	ز
r per *w/ū*	ر	و
alif per *l* (più raramente)	ا	ل

Questo errore nell'individuare le forme dei caratteri non include l'elisione di punti diacritici, a meno che non intervengano ulteriori decorazioni esterne alla parola o deterioramenti dell'immagine. Il riconoscimento erroneo delle forme ha effetto sulla composizione del grafema; come anticipato, un grafema come د <*d*> rappresentato con la sua base sotto il rigo potrebbe condurre al riconoscimento di un ر <*r*> e, viceversa, <r> in cui la parte finale si alzi sopra al rigo potrebbe generare una confusione grafica per un sistema OCR. Lo stesso può avvenire tra <*ḏ*> e <*z*>. L'ultimo esempio riportato nella tabella (*alif* per *lām*) è assai meno frequente e spesso condizionato da variabili "esterne" che possono essere di diversa natura. Nel caso particolare riportato in Fig. 1 la mancanza di inchiostro nella legatura di *lām* sulla matrice di stampa comporta il riconoscimento del grafema *lām* come *alif*.

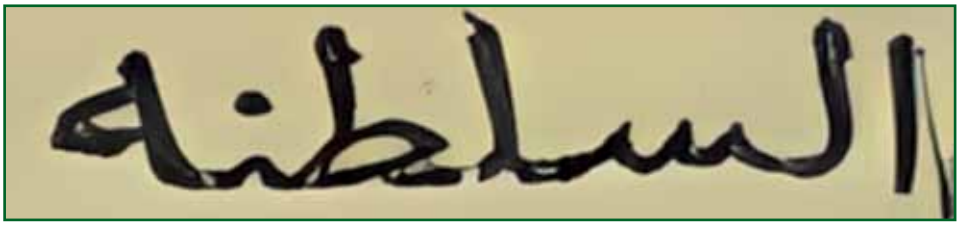

Figura 1. *Esempio di distanziamento tra i caratteri tra lām e carattere successivo nella parola "al-sulṭana"*

Scambio di forme non simili (con e senza diacritici)

Cout per Cinp	**Cout**	**Cinp**
h per *m* (in posizione iniziale)	هـ	مـ
m per *h* (in posizione iniziale)	مـ	هـ
k per *h* (solo in alcuni casi, si veda sotto)	كـ	هـ
ṯ per *š* (non in posizione finale, più raramente in posizione iniziale)	ث	ش

Se si prendono come esempio i grafemi ث e ش, può accadere che assumano forme simili quando i denti tipici di <*š*> sono appiattiti sul rigo (in stili come *ruqʿa*, *taʿlīq* e *nastaʿlīq*). L'errore di scambio di forma è soggetto quindi alle variabili definite qui come "interne" e proprie della calligrafia araba. L'esempio di <*h*> per <*k*> sembra pertanto graficamente il più indicativo per la suddivisione qui proposta. Lo scambio di <h> per <k> si può riscontrare soprattutto quando la ك <*k*> espressa in stili quali *ruqʿa* e *nastaʿlīq* nella posizione iniziale e mediana vede la sua parte più prossima al rigo arrotondarsi a formare un occhiello. Quando rappresentata in questa forma può essere erroneamente riconosciuta come هـ <*h*> (si veda Fig. 2).
Le variabili "interne" allo stile calligrafico sono anche responsabili della fusione dei punti diacritici; due punti diacritici allineati possono infatti essere orientati in modo diverso e anche fusi insieme in un unico tratto parallelo al rigo. In casi particolari i punti diacritici possono essere disposti anche in senso verticale, aumentando la possibilità di errore. Caratteri calligrafici come il *ruqʿa* esprimono i due punti, sopra o sotto al rigo, come un trattino orizzontale e i tre punti come una sorta di accento circonflesso.

Figura 2. *Esempio dello stile ruqʿa con punti diacritici uniti in trattino nella parola "al-muškilāt"*

Rimozione totale o parziale dei punti diacritici

Cout per Cinp	**Cout**	**Cinp**
n per *t* (non in posizione finale)	ن	ت
t per *ṯ*	ت	ث
ḥ per *ẖ*	ح	خ
f per *q*	ف	ق

La rimozione dei punti diacritici può essere totale o parziale, causando convergenza verso un grafema diverso (خ <*ẖ*> in ح <*ḥ*>) o forme non grammaticali. Si prenda ad esempio in considerazione il termine *taẖayyur* تحير («scegliere, optare») dalla radice خ-ي-ر. In questa forma la parola, priva di vocalizzazione e contesto, può rappresentare sei parole distinte (non omofone), di cui cinque verbi (tre con diatesi attiva e due con diatesi passiva) e un sostantivo. Al contempo, la parola può appartenere a classi grammaticali diverse e avere significati diversi pur rimanendo graficamente invariata. Ponendo ora che il riconoscimento OCR abbia erroneamente eliso il punto diacritico del grafema خ generando ciò che segue: تحير ("confusione, l'essere confuso"), è dunque evidente

che un errore di riconoscimento può avere conseguenze non limitate al contesto grafico ma anche sul piano semantico, sebbene riconosciuta da una risorsa linguistica come parola valida in un'ontologia o un modello linguistico. L'esempio vuole rimarcare le dimensioni del problema, che rappresenta un concreto ostacolo all'applicazione a posteriori di strumenti di NLP (come ad esempio uno *stemmer*[26]), comportando il rischio di propagare l'errore nelle fasi successive. Dunque l'errore inizialmente grafico genera conseguenze semantiche di non poco conto.

Nel contesto dell'ideazione e progettazione di un flusso di lavoro catalografico, errori sui diacritici dei nomi d'autore possono generare forme ambigue, aumentando i tempi della fase di verifica e controllo catalografico. La precisione nell'estrazione dati dai frontespizi è cruciale per evitare tali problematiche.

A tal proposito un esempio è proposto da Sublet[27] nel contesto degli studi onomastici arabi in riferimento all'opera di al-Ḏahabī dal titolo *al*-Muštabih fī al-riğāl[28], nel quale sono riportati i possibili esiti di lettura di elementi onomastici omografi[29] con variazioni di punti e segni diacritici: الجُندي، الجَندي or الحيدي . Al-Ḏahabī specifica anche nel dettaglio lo scambio di forme simili riportando di nuovo gli esiti di lettura possibili, come ad esempio: الحيري الجَنْزي، الخَبْري، الخَبَري، الجبْري، الحُتْري، الجَبْري، .

Gli esempi riguardanti ه/ة e ي/ى, non riportati nella tabella per brevità, sono ben noti nella linguistica computazionale e negli studi di ANLP (Arabic Natural Language Processing), in particolare nell'ambito dello sviluppo di sistemi di segmentazione e di NLP per la lingua araba. Proponendo un segmentatore per l'arabo classico, Mohamed e Sayyed[30] categorizzano questi due gruppi come "t/h confusion set" e "y confusion set" da considerarsi a fianco del trattamento della *alif-hamza*, "the hamza confusion set", nel contesto dell'ortografia araba standard o sub-standard[31].

Aggiunta di punti diacritici

In maniera inversa alla rimozione dei punti diacritici, alcune variabili "esterne" (decorazioni sul testo, deterioramento dell'immagine o della stampa), come "interne" ("rumore" generato da compresenza di vocalizzazione e punti diacritici), possono concorrere, nell'output, all'aggiunta di punti diacritici ove non dovrebbero essere presenti. L'"errore" in questione è diametralmente opposto al precedente ma ne condivide numerose caratteristiche con conseguenze simili a quelle descritte nel punto precedente. Per questa ragione non verranno trattate nello specifico.

---

[26] Gli algoritmi di normalizzazione post-processing rimuovono diacritici, articoli, prefissi e suffissi. Tra i principali *stemmer*: Buckwalter (2002), Tashaphyne Light Stemmer, CAMeL lab (Obeid et al. 2020)

[27] Sublet 1999, p. 125.

[28] Ḏahabī 1962.

[29] In questo caso la *nisba*, un elemento onomastico costituito dall'aggiunta di un suffisso derivativo *-yā' (o -yā' nisbiyya* ـيّ) a un sostantivo. In alcuni casi, la *-yā›* può essere preceduta (per ragioni fonetiche) dal nesso *alif-nūn* (انـي = ي + ان) o succeduto da *tā' marbūṭa* (ـيّة) qualora femminile.

[30] Mohamed — Sayyed 2019, p. 4-5.

[31] Data il numero consistente di rappresentazioni dei grafemi arabi e persiani, anche in ambito computazionale sono numerosi gli studi sull'ortografia. Si veda: (Habash 2010, p. 31-34; Habash — Diab — Rambow 2012, p. 711-718).

Scivolamento dei punti diacritici

Cout	Cinp
تقتيش	تقتيش
محبور	مجبور
تحميس	تخميس
تناقد	تناقد

Uno dei più comuni "errori" è quello generato dallo spostamento dei diacritici, collegato alle variabili "interne" dello stile calligrafico. Questo tipo di "errore" è soggetto anche al grado di addestramento e taratura del sistema OCR sulla distinzione dei caratteri e dei collegamenti fra di essi. La sovrapposizione dei caratteri attraverso specifiche legature sposta i punti diacritici in posizioni che possono risultare ambigue (Fig. 3) per un sistema non sufficientemente addestrato. La posizione e la forma assunta dai grafemi dovuti al tipo di legatura può inoltre prevedere un posizionamento non necessariamente parallelo al rigo.

Figura 3. *Esempio di legatura per la parola "naǧm": a destra con legatura distesa, a sinistra contratta*

Se si prende in considerazione lo stile calligrafico *nastaʿlīq* nella sua varietà *šekasteh*[32] (pers.: "rotto, spezzato"), l'orientamento del rigo, inclinato a 45 gradi, in aggiunta a un set di legature specifiche, rende il riconoscimento arduo per sistemi OCR non sufficientemente sviluppati. Un esempio concreto è fornito in Fig. 4, in corrispondenza del nesso consonantico <*n-s-t*> (نستـــ), dove il punto diacritico relativo a ن <*n*> è slittato verso il centro della stringa, i tre "denti" di س <*s*> sono appiattiti sul rigo mentre i due punti di ت <*t*> sono spostati verso l'inizio della stringa. Dal punto di vista delle conseguenze sul riconoscimento si può notare che lo spostamento dei diacritici può comportare l'aggiunta o la sottrazione dei segni diacritici, con esiti del tutto assimilabili a quelli già descritti.

[32] Nella variante *nastaʿlīq-šekasteh*, ogni elemento influenza il successivo trovando un'unità grafica nel suo risultato finale e sviluppandosi in modo inclinato sul rigo. Grafemi, parole e legature sono continue e formano l'una la base per la successiva, seguendo una forma ellittica (Maʿnavī Rād 2013, p. 23).

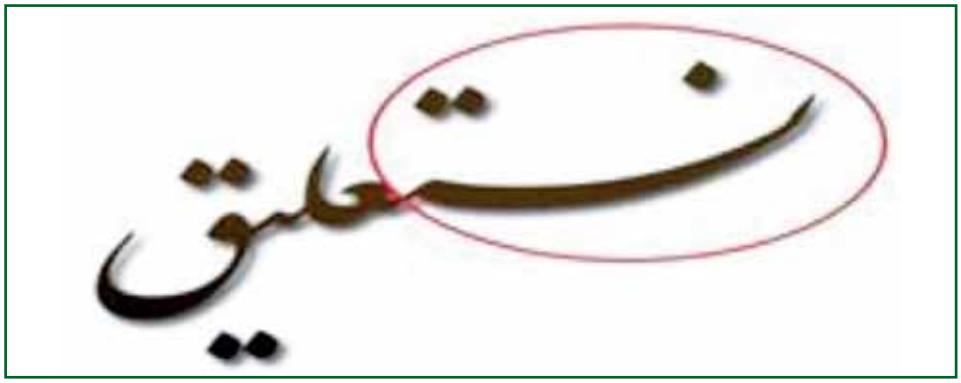

Figura 4. *Cerchiato in rosso si trova il nesso consonantico n-s-t rappresentato con lo stile šekasteh per la parola "nastaʿlīq"*

Interferenza rumorosa tra segni e punti diacritici

Cout per Cinp	**Cout**	**Cinp**
ġ per *ʿu*	غذير	عُذير
t per *nu*	تور	نُور
z per *ra*	ضَرْائب	ائبُضَر
[33]*ġ* per *ʿ (sukūn)*	فعْل	فعُل

Nel corso dei test si sono potute rilevare interferenze tra vocali e punti diacritici, risultanti in un riconoscimento erroneo di punti diacritici. Tuttavia, se è vero che le vocali possono essere rese erroneamente come punti diacritici, non è altrettanto vero il contrario. Infatti, come già rilevato, generalmente i punti diacritici posso subire trattamenti scorretti, come già indicato in precedenza, possono essere rimossi, aggiunti, dimenticati o traslati orizzontalmente, o in alcuni casi verticalmente. Quando presenti, le vocali possono interferire con i punti diacritici, causandone spesso l'aggiunta dove non sono previsti. Questa interferenza ha come conseguenza lo scambio delle consonanti, generando forme non grammaticali o lemmi di significato diverso da quello della parola in input.

Riconoscimento dovuto a dimensioni dei caratteri

Talvolta la variazione della dimensione del carattere può essere decisiva nel determinare la buona riuscita di un riconoscimento, concorrendo con altre variabili sia "interne" che "esterne" a restituire un esito "errato". Poniamo qui l'esempio di due parole con il medesimo schema morfologico *fiʿāla* (plur. *fiʿālāt*), dal momento che le vocali lunghe sono due *alif* generalmente estratte correttamente. Ciò che si modifica sono le consonanti che la compongono ma con inaspettate "evoluzioni" grafiche:

[33] Segno diacritico apposto su consonanti per indicare l'assenza di vocale in lettura, *sukūn* "quiete, assenza di movimento, tranquillità"

MODUS

Cout per Cinp	Cout	Cinp
	صراعات	دراسات
ṣ per *d*	صـ	د
ʿ per *s*	عـ	سـ

Il primo termine a destra, *dirāsāt* ("studi"), rappresenta *l'input* mentre il termine a sinistra, *ṣirāʿāt* ("lotte, conflitti"), la stringa effettivamente generata dall'OCR. Se si analizzano attentamente i caratteri "errati", ad esempio nella riga ع <ʿ> - س <s>, l'"errore" appare poco comprensibile. I due grafemi hanno due forme sopra il rigo con un basso grado di similarità. Quando l'*occhiello* concavo con apertura a destra tipico della <ʿ> (عـ) si trova sopra al rigo, differisce molto dal punto di vista grafico rispetto ai tre *denti* della <s> (سـ). Nonostante le palesi differenze formali, però, i due caratteri vengono confusi. Questo esito suggerisce alcune interessanti osservazioni: da un lato si può notare come la dimensione influenzi il riconoscimento anche in casi macroscopici, qualora il sistema OCR non sia in grado di garantire una adeguata robustezza (variabile "interna"); dall'altra la dimensione ridotta comporta una compressione dei grafemi e un aumento della probabilità che si verifichi uno scambio di forma soprattutto in prossimità del rigo[34].

Risultati derivanti da trattamento di *alif* e *hamza*

Cout per Cinp	Cout	Cinp
alif per *alif-hamza*	قول	أقول
	سأل	سأل
alif maqṣūra per *alif-hamza* (in contesto di vocale *i)*	قرى	قرئ
punto diacritco per *hamza* (e viceversa, sotto)	سذال	سؤال

Il trattamento del nesso *alif-hamza* rappresenta una delle difficoltà che la grafia araba pone, non solo per il riconoscimento ottico dei caratteri ma anche per le tecniche di analisi testuale (analizzatori morfologici, *stemmer*, *tagger* ecc.) prodotte nel campo dell'ANLP[35]. A tal proposito si prenda in considerazione il quarto esempio سـأل/سال , dove la lettura della *hamza* è fondamentale per distinguere il verbo *saʾalᵃ* "chiedere, domandare" (media *hamza*) e il verbo *sālᵃ* "scorrere, fluire, divenire/essere liquido". In questo caso la ء distingue due radici afferenti a due campi semantici ben distinti: *s-ʾ-l* e *s-y-l*. Analogo è anche l'esempio del trattamento della ء sostenuto da *alif maqṣūra*

[34] I test condotti hanno sottolineato che con immagini di buona qualità, anche con caratteri di piccola dimensione, i grafemi sopra il rigo (alif ا) e isolati (come t in -āt ات del plurale femminile) vengono riconosciuti correttamente.

[35] Il trattamento della *hamza* e delle varianti della *alif* è stato affrontato in ambito ANLP (morfologia, stemming, tagging, tokenizzazione, lemmatizzazione) e, negli studi OCR, dal punto di vista della segmentazione grafica dei caratteri. Si veda ad esempio (Qaroush et al. 2020, p. 2-5).

(ى) qui riportato nella I forma del verbo passivo III p.s. della radice <*q-r-ʾ*> ("leggere, recitare, salmodiare, studiare"). Nuovamente, la forma passiva قرئ *quriʾa* si distingue dal plurale fratto قرى *qurā* ("villaggi") solamente per la presenza della *hamza*. Per questo motivo non si può prescindere, dal punto di vista dell'analisi della semantica, da risorse linguistiche capaci di disambiguare le parole.

Mancata lettura degli spazi

Il rilevamento degli spazi è uno dei temi più studiati nel campo dell'ANLP, in particolare in relazione alla normalizzazione e alla segmentazione del testo. In quest'analisi, tuttavia, il testo è considerato come unità da riportare semplicemente nell'output in maniera più fedele possibile a ciò che si trova sul foglio oggetto dell'analisi. Qui di seguito si riportano due esempi:

Cout	Cinp
أحمد بنعلي	أحمد بن علي
وَرَبَّه	وَرَبُّه

La mancata lettura degli spazi può generare forme inesistenti o semplicemente diverse dall'input in maniera del tutto simile alle ambiguità generate con sottrazione, addizione e scivolamento di diacritici. Nel secondo esempio, la *wa* di congiunzione in arabo è prefissa alla parola che segue e nel primo nesso a destra si ha *wa-rabbᵘhᵘ* ("e il suo signore, padrone") mentre nel nesso più a sinistra si può considerare *wa* come parte della radice *w-r-b* "essere marcio, marcire" alla II forma *warrabᵃ-yuwarribᵘ* coniugata alla III p.s.m passato con suffisso *-hu* ("egli, lui"), quindi *warraba-hu*: letteralmente "lo espresse in maniera contraddittoria". Si tratta di un esempio limite che esemplifica chiaramente il modo in cui aspetti grafici si intersecano con quelli semantici e morfologico-grammaticali.

Risultati con "errori" multipli (o composti)

Un'ulteriore categoria è rappresentata dagli errori multipli (o composti), ovvero di diversa tipologia occorrenti nel risultato di riconoscimento della medesima stringa di grafemi o di testo. Qui di seguito alcuni esempi esplicativi di questa categoria:

Cout	Cinp
مشود نزويگ	ميشود نزديک
تژس ولذذ	ترس ولرز
النبوي صحعحلفمهر	النبوي القصص صحيح

Anche in questo caso, bisogna sempre tenere in considerazione la presenza di una commistione di variabili "interne" ed "esterne" che concorrono nel generare errori. Sono possibili combinazioni quali: scambio di forme, rimozione e addizione di diacritici, mancata lettura di spazi e interferenze tra vocalizzazione e punti diacritici. Estendendo le stringhe testuali e passando da unità minime (glifi e grafemi) a unità testuali più com-

plesse (parole, frasi, paragrafi ecc.) diviene più facile trovare una quantità maggiore di errori e, in caso di OCR non sufficientemente addestrati, errori multipli di crescente complessità fino a casi di stringhe completamente irriconoscibili.

Segni di lettura e diacritici nel Corano

Per quanto concerne il Corano, sono presenti online numerosi strumenti per l'analisi del testo, come anche versioni già digitalizzate e navigabili[36]. Tuttavia, nel caso di altri testi non digitalizzati in cui si trovano riferimenti diretti ai versetti coranici (*tafsīr*, *fiqh*, raccolte di *ḥadīṯ* ecc.) le stringhe di testo possono presentare una serie supplementare di segni diacritici sopra i grafemi, per segnalarne la corretta recitazione. In questo caso nel testo non vocalizzato sono inseriti i versetti coranici a cui si fa riferimento completi di vocalizzazione e segni diacritici come nella Fig. 5.

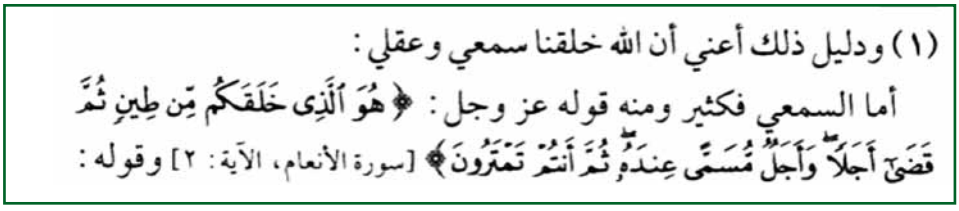
(١) ودليل ذلك أعني أن الله خلقنا سمعي وعقلي:
أما السمعي فكثير ومنه قوله عز وجل: ﴿هُوَ ٱلَّذِى خَلَقَكُم مِّن طِينٍ ثُمَّ
قَضَىٰٓ أَجَلًا ۖ وَأَجَلٌ مُّسَمًّى عِندَهُۥ ثُمَّ أَنتُمْ تَمْتَرُونَ﴾ [سورة الأنعام، الآية: ٢] وقوله:

Figura 5. *Frammento di testo da al-ʿUṯaymīn, M. 2004. Šarḥ ṯalāṯ al-ʾuṣūl, Riyāḍ, dār al-ṯurayyā li-l-našr, p. 29*

Questi particolari segni diacritici sono l'espressione grafica delle minuziose regole di dizione e lettura coranica che costituiscono la scienza islamica del *taǧwīd*[37]. I segni diacritici relativi alle vocali, definiti nel *taǧwīd* come segni di "precisazione" (*ʿalāmāt al-ẓabṭ*)[38] possono eventualmente aumentare la quantità di "rumore" sul testo. A titolo di esempio si vogliono ricordare: la piccola *m* sovrascritta in contesto di *n-b* o *m-b* (*iqlāb*); il simbolo che indica la presenza di *imāla*[39] o di *išmām*[40], che consiste in un circoletto o rombo pieno posizionato sotto ◌ٖ o sopra il rigo ◌ٗ (*dāʾira maṭmūsa*) rispettivamente, es. Cor. XI:41 مَجْرٖاهَا (*maǧrā-hā*) e in Cor. XII:11 تَأْمَنُنَا (*taʾmannā*); o ancora i segni di allungamento (*madd*) — come quello riscontrabile sulla *alif-madda* anche al di fuori del testo coranico. La *alif* sovrascritta (detta anche *alif ḫanǧariyya*) stante alla presenza di una *alif* di prolungamento non graficamente espressa nella sua forma consueta, come in Q. CXIII:4 (نفٰثٰت).

[36] Si veda ad esempio il progetto open source The Quranic Arabic Corpus: <https://corpus.quran.com/> e il Tanzil Project: <https://tanzil.net/>.

[37] *taǧwīd* significa letteralmente: "articolare i fonemi nel loro punto appropriato rispettando tutte le caratteristiche" (Zalaṭ bin Rafʿat 2006, p. 5).

[38] Šukrī 2014, p. 1496.

[39] *ʿimāla* (letteralmente "l'inclinazione, il tendere verso qc."), in fonetica araba è usato ad indicare l'innalzamento della vocale a verso la vocale *i* causata dall'influenza della *kasra* sulla *alif* stante per vocale a. Tale fenomeno era già stato osservata dal famoso Sībawayhi (1988, p. 259-261).

[40] Con *išmām* si intende il posizionamento delle labbra per produrre una vocale *u, ḍamma* (o vocale i, *kasra*) ma senza emissione di suono effettivo (Naṣr 2009, p. 615).

MODUS

Conclusioni e prospettive di ricerca

Questo studio vuole mettere in evidenza alcuni dei comportamenti comuni dei sistemi OCR open-source testati su frontespizi arabi, analizzando le relazioni tra questi sistemi e le peculiarità dell'alfabeto arabo nel contesto di in un possibile flusso di lavoro di catalogazione automatizzata. Il riconoscimento ottico dell'arabo rimane un campo in parte inesplorato, sia dal punto di vista teorico che pratico. Gli "errori" identificati possono essere causati da fattori intrinseci all'alfabeto o "esterni", che si riducono spesso alla rimozione o aggiunta di punti diacritici. I risultati teorici trovano riscontro in ricerche analoghe, come l'analisi di Kiessling su KrakenOCR[41], che evidenzia problemi simili: lettura degli spazi, difficoltà con diacritici, trattamento del nesso *alif-hamza* e riconoscimento di consonanti specifiche[42].

Uno dei problemi principali dell'AOCR (OCR di caratteri arabi) è la scarsità di set di dati disponibili rispetto ad altri alfabeti[43], in particolare per quanto riguarda i testi definiti nel contesto OCR come "historical". Nonostante la crescita di database annotati, mancano ancora *corpora* storici di alta qualità per sistemi multilingua. Nel caso specifico del progetto Digital Maktaba si registra invece la necessità, a cui si sta lavorando attualmente, di creare un set di dati specifico per frontespizi a stampa.

Un altro importante tema affrontato nel corso di questo studio riguarda la codifica Unicode e, in generale, la mappatura dei caratteri. La combinazione OCR-Unicode è fondamentale per risultati ottimali. Lo standard Unicode presenta ambiguità di codificazione e limitatezze per gli stili calligrafici meno rappresentati. Ad esempio, il carattere ى ha due codici (U0649 arabo, U06CC persiano) per la stessa rappresentazione grafica. Queste ambiguità, dovute all'omografia araba, hanno ripercussioni sull'individuazione di informazioni ai fini catalografici. I numerali presentano set differenti: arabo (0660-0669) e persiano/altre lingue (06F0-06F9), tema finora poco trattato ma rilevante per il recupero dati catalografici.

Interessanti studi come quello di Milo e Martínez[44] hanno introdotto il concetto di *arcigrafema*, evidenziando la necessità di un addestramento linguistico per OCR dinamici con intelligenze artificiali. Allo stesso modo considerazioni sull'ambiguità della scrittura araba come un "meccanismo di tolleranza", evidente nei manoscritti coranici con doppi marcatori diacritici[45] sono fondamentali per avanzare il dibattito anche nel campo OCR all'interno del contesto condiviso delle *Digital Humanities*.

Al contempo non si devono sottovalutare le possibilità tecniche derivanti dalle possibilità offerte dai *Transformers* (che elaborano sequenze di dati tramite un meccanismo che valuta l'importanza relativa di ogni elemento rispetto agli altri) e dai recenti sviluppi nel campo dei modelli linguistici. Lo sviluppo di algoritmi *Deep Learning* (reti neurali artificiali multistrato) ha rinnovato l'interesse per le lingue araba, persiana, curda e

[41] Romanov et al. 2017.
[42] Kiessling et al. 2021, p. 10-11.
[43] Wagaa — Kallel — Mellouli 2022, p. 1.
[44] Milo — Martínez 2019.
[45] Fedeli 2020.

urdu. I modelli di distribuzione basati su *Vector Semantics* (semantica distribuzionale[46]) potrebbero migliorare l'analisi contestuale tramite disambiguazione semantica[47].
Modelli come BERT, AraBERT e LLMs (come Mistral, LaMa, o GPT) offrono prospettive per il NLP arabo attraverso *word embeddings* (rappresentazione distribuita delle parole) e previsione contestuale nella fase di post-elaborazione. L'integrazione fra OCR e LLM rimane però poco esplorata per la scrittura araba, nonostante applicazioni promettenti nell'estrazione di informazioni[48].
Infine, risultano di un certo interesse per future prospettive di ricerca le analogie riscontrate con alcuni studi cognitivi sulla lettura: i sistemi OCR potrebbero infatti essere paragonati al processo di apprendimento della lettura infantile, sostituendo conoscenze pregresse con risorse linguistiche. Gli studi di Azzam e Abu-Rabia per l'arabo[49] e Baluch per il persiano[50] potrebbero fornire spunti preziosi per l'analisi degli errori OCR confrontandoli con quelli dei giovani lettori.

In digitisation in the context of document preservation in large archives, OCR systems have become essential tools as a compromise between good character recognition and low cost. However, they are still lacking in training on historical and religious texts and in particular with non-Latin alphabets. Large amounts of unstructured data characterised by scatter and noise expose the limitations of text mining techniques on several levels. These limitations add up to contextual variables hindering OCR systems from achieving optimal character recognition. This represents a real problem when considering such systems as central to the development of applications involving the downstream use of other NLP techniques. From this point of view, an error analysis is part of a post-processing phase that can have corrective effects on the output in order to improve the recognition result especially when combined with a context analysis. Greater attention to post-processing on both glyphs and graphemes could bring about a considerable improvement in OCR effectiveness by significantly advancing the current state of the art. The brief study presented here aims to outline common traits of errors committed by such OCRs tested in the Digital Maktaba project.

[46] Sadrzadeh — Muskens 2018.
[47] Gavin 2018, p. 660-663.
[48] Tang et al. 2024; Xu et al. 2023.
[49] Azzam 1989, Abu-Rabia 1998.
[50] Baluch 2005.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

Riferimenti bibliografici

Abu-Rabia 1998 — Salim Abu-Rabia. *Reading Arabic Texts: Effects of Text Type, Reader Type and Vowelization*. «Reading and Writing», 10 (1998), n. 2, p. 105–119. <https://doi.org/10.1023/A:1007906222227>.

Afli — Loïc —Schwenk 2016 — Haithem Afli — Barrault Loïc — Holger Schwenk. *OCR Error Correction Using Statistical Machine Translation*. «International Journal of Computational Linguistics and Applications», 7 (2016), n. 1, p. 175–191.

Ahmed — Abidi 2019 — Muna Ahmed — Ali Abidi. *Review on Optical Character Recognition*. «International Research Journal of Engineering and Technology (IRJET)», 6 (2019), n. 6, p. 3666–3669.

Alaasam — Barakat — Kassis — El-Sana 2017 — Reem Alaasam — Berat Barakat — Majeed Kassis — Jihad El-Sana. *Experiment study on utilizing convolutional neural networks to recognize historical Arabic handwritten text.* In: *1st International Workshop on Arabic Script Analysis and Recognition* (ASAR), 2017, p. 124–128. <https://doi.org/10.1109/ASAR.2017.8067773>.

Alghamdi — Alkhazi — Teahan 2016 — Mansoor A. Alghamdi — Ibrahim S. Alkhazi — William J. Teahan. *Arabic OCR evaluation tool.* In: *2016 7th International Conference on Computer Science and Information Technology* (CSIT), 2016, p. 1–6. <https://doi.org/10.1109/CSIT.2016.7549460>.

Azzam 1989 — Rima Azzam. *Orthography and Reading of the Arabic Language.* In: *Reading and Writing Disorders in Different Orthographic Systems*, ed. by P. G. Aaron, R. Malatesha Joshi (NATO ASI Series; 52). Dordrecht: Springer Netherlands, 1989, p. 203–218. <https://doi.org/10.1007/978-94-009-1041-6_12>.

Baluch 2005 — Bahman Baluch. *Persian Orthography and Its Relation to Literacy.* In: *Handbook of Orthography and Literacy*, ed. by R. Malatesha Joshi, P. G. Aaron. London: Lawrence Erlbaum Associates, 2005, p. 365-376. <https://eprints.mdx.ac.uk/7649/>.

Bassil — Alwani 2012 — Youssef Bassil — Mohammad Alwani. *OCR Post-Processing Error Correction Algorithm using Google Online Spelling Suggestion*. 2012. <https://arxiv.org/abs/1204.0191>

Batawi — Abulnaja 2012 — Y. Batawi — Osama Abulnaja. *Accuracy Evaluation of Arabic Optical Character Recognition Voting Technique: Experimental Study*. «International Journal of Electrical & Computer Sciences IJECS-IJENS», 12 (febbraio 2012), p. 29–33.

Bergamaschiet et al. 2022 — Sonia Bergamaschi — Stefania De Nardis — Riccardo Martoglia — Federico Ruozzi — Luca Sala — Matteo Vanzini — Riccardo Amerigo Vigliermo. *Novel Perspectives for the Management of Multilingual and Multialphabetic Heritages through Automatic Knowledge Extraction: The DigitalMaktaba Approach*. «Sensors», 22 (2022), n. 11, 3995. <https://doi.org/10.3390/s22113995>.

Boros et al. 2024 — Emanuela Boros — Maud Ehrmann —Matteo Romanello — Sven Najem-Meyer — Frédéric Kaplan. *Post-correction of Historical Text Transcripts with Large Language Models: An Exploratory Study*. In: *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*. St Julian (Malta): The Association for Computational Linguistics, 2024, p. 133-159.

Buckwalter 2002 — Tim Buckwalter. *Buckwalter Arabic Morphological Analyzer Version 1.0*. Linguistic Data Consortium, 2002. <https://doi.org/10.35111/7VZM-MB15>.

Buziyād 2011 — Sāmiya Buziyād. *mağallat al-dirāsāt al-luġawiyya*. «ẓāhirat al-rawm wa-l-'išmām bayna al-nuḥḥāt wa-l-qurrā': dirāsa ṣawṭiyya», 13 (2011) p. 65–84.

Chiron eet al. 2017 — Guillaume Chiron — Antoine Doucet — Mickael Coustaty — Muriel Visani — Jean-Philippe Moreux. *Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information*. In: *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2017, p. 1–4. <https://doi.org/10.1109/JCDL.2017.7991582>.

Ḏahabī 1962 — Šams al-dīn al-Ḏahabī. *al-Muštabih fī al-riǧāl: asmā'i-him wa ansābi-him*, a cura di 'Alī Muḥammad al-Bağāwī. Bayrūt: Dār iḥiyā' al-kutub al-'arabiyya, 1962.

Drobac — Lindén 2020 — Senka Drobac — Krister Lindén. *Optical character recognition with neural networks and post-correction with finite state methods*. «International Journal on Document Analysis and Recognition (IJDAR)», 23 (2020), p. 279–295. <https://doi.org/10.1007/s10032-020-00359-9>.

Fedeli 2020 — Alba Fedeli. *The Qur'ānic Text from Manuscript to Digital Form: Metalinguistic Markup of Scribes and Editors*. In: *From Scrolls to Scrolling, Sacred Texts, Materiality, and Dynamic Media Cultures*., ed. by B. A. Anderson. De Gruyter, 2020, p. 213–246. <https://doi.org/10.1515/9783110634440-010>.

MODUS

Garbini — Durand 1994 — Giovanni Garbini — Olivier Durand. *Introduzione alle lingue semitiche*. (Studi sul vicino Oriente antico; 2). Brescia: Paideia, 1994.

Gavin 2018 — Michael Gavin. *Vector Semantics, William Empson, and the Study of Ambiguity*. «Critical Inquiry», 44 (2018), n. 4, p. 641–673. <https://doi.org/10.1086/698174>.

Ğubūrī 1998 — Muhammad Šukr al-Ğubūrī. *al-ḫaṭṭ al-ʿarabī wa-l-zaḫrafa al-ʻislāmiyya*. Irbid, al-ʻUrdun: Dār al-ʻamal li-l-našr wa-l-tawzīʿ, 1998.

Habash 2010 — Nizar Habash. I*ntroduction to Arabic Natural Language Processin*g. Cham: Springer, 2010. <https://link.springer.com/book/10.1007/978-3-031-02139-8>.

Habash — Diab — Rambow 2012 — Nizar Habash — Mona Diab — Owen Rambow. *Conventional Orthography for Dialectal Arabic*. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012, Istanbul*). Paris: European Language Resources Association, 2012, p. 711-718.

Hajiali 2023 — Mahdi Hajiali. *OCR Post-Processing Using Large Language Models*. (Dissertation). Las Vegas: UNLV, 2023. <https://digitalscholarship.unlv.edu/thesesdissertations/4811>.

Ibn-Durustawaih et al. 1998 — Ibn-Durustawaih — Abdallah Ibn-Gafar — Muhammad Badawi al-Mahtun — Ramadan Abd-at-Tauwab. *Tashih al-Fasih wa-sarhihi*. al-Qahira: Gumhuriyat Misr al-`Arabiya, Wizarat al-Auqaf, al-Maglis al-A`la li-s-Suun al-Islamiya, 1998.

ISIRI 6219 2002 — Institute of Standards and Industrial Research of Iran. *ISIRI 6219 - FarsiWeb*. 2002. <https://persian-computing.org/archives/Sharif-FarsiWeb-Inc/ISIRI_6219.html>.

Javed — Hussain 2009 — Sobia Tariq Javed — Sarmad Hussain. *Improving Nastalique specific pre-recognition process for Urdu OCR*. In: *IEEE 13th International Multitopic Conference (Islamabad, 2009)*. Piscataway (NJ): IEEE, 2009, p. 1–6. <https://doi.org/10.1109/INMIC.2009.5383111>.

Kanungo — Marton — Bulbul 1999 — Tapas Kanungo — Gregory A. Marton — Osama Bulbul. *Performance evaluation of two Arabic OCR products*. In: *27th AIPR Workshop: Advances in Computer-Assisted Recognition (29 January 1999)*. SPIE, 1999, p. 76-83. <https://doi.org/10.1117/12.339809>.

Kew 2005 — Jhonathan Kew. *Notes on Some Unicode Arabic Characters: Recommendations for Usage*. Draft 2. 21 aprile 2005. <https://it.scribd.com/doc/235215269/Arabic-Letter-Usage-Notes>.

Khosrobeygi et al. 2020 — Zohreh Khosrobeygi — Hadi Veisi — Hamid Reza Ahmadi — Hanieh Shabanian. *A rule-based post-processing approach to improve Persian OCR performance*. «Scientia Iranica», 27 (2020), n. 6, p. 3019–3033. <https://doi.org/10.24200/sci.2020.53435.3267>.

Kiessling et al. 2021 — Benjamin Kiessling — Gennady Kurin — Matthew Thomas Miller — Kader Smail. *Advances and Limitations in Open Source Arabic-Script OCR: A Case Study*. «Digital Studies / Le Champ Numérique», 11 (2021), n. 1. <https://doi.org/10.16995/dscn.8094>.

Lībī 2017 — A. al-Mālikī al-Lībī. *al-tuḥfa al-mālikiyya fī talḫīṣ 'uṣūl riwāya ḥafṣ ʿan ʿāṣim min ṭarīq al-šāṭibiyya*. 1a ed. silsila mulaḫḫaṣāt 'uṣūl al-qirā'āt 2., 2017.

Maʿnavī Rād 2013 — Mītrā Maʿnavī Rād. *Negareh: faṣlnāmeh ʿelmī – pajūheši-ye negareh. In: taʿāmol-e sāḫtār va sabok dar šekasteh nevīsī: šafīʿān, darvīš va golestāneh*. 27 (2013), p. 21–33.

Milo 2006 — Thomas Milo. *The original Tasmeem Manual*. 2006. <https://www.academia.edu/3517400/The_original_Tasmeem_Manual_2006_>.

Milo — Martínez 2019 — Thomas Milo — Alicia González Martínez. *A New Strategy for Arabic OCR: Archigraphemes, Letter Blocks, Script Grammar, and shape synthesis*. In: *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage (DATeCH2019)*. New York: Association for Computing Machinery, 2019, p. 93–96. <https://doi.org/10.1145/3322905.3322928>.

Mohamed — Sayyed 2019 — Emad Mohamed — Zeeshan Ali Sayyed. 2019. *Arabic-SOS: Segmentation, Stemming, and Orthography Standardization for Classical and pre-Modern Standard Arabic*. In: *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage (DATeCH2019)*. New York: Association for Computing Machinery, 2019, p. 27–32. <https://doi.org/10.1145/3322905.3322927>.

Munağğid 1999 — Muḥammad Nūr al-dīn al-Munağğid. *al-ištirāk lafẓī fi-l-qur'ān al-karīm: bayna al-naẓariyya wa-l-taṭbīq. Dimašq: Dār al-fikr*, 1999. <https://ebook.univeyes.com/95981/pdf-الاشتراك اللفظي في القرآن الكريم بين النظرية والتطبيق>.

MODUS

Naṣr 2009 — Ismāʿīl bin Ḥamād al-Ǧawharī abū Naṣr. *al-ṣaḥḥāḥ*. al-Qāhira: Dār al-ḥadīṯ, 2009.

Naṣrāwī 2015 — ʿādil ʿabbās al-Naṣrāwī. *dawāt / mağallah faṣliyyah tuʿnà bi-l-buḥūṯ wa-l-dirasāt al-luġawiyyah*. «al-manhağ al-riyāḍī fī al-dars al-muʿğamīʿind al-Farāhīdī: taqwīm wa taǧdīd», 1 (2015), n. 3, p. 75–101.

Nguyen et al. 2021 — Thi Tuyet Hai Nguyen — Adam Jatowt — Mickael Coustaty — Antoine Doucet. *Survey of Post-OCR Processing Approaches*. «ACM Computing Surveys», 54 (2021), n. 6, p. 1–37. <https://doi.org/10.1145/3453476>.

Obeid et al. 2020 — Ossama Obeid — Nasser Zalmout — Salam Khalifa — Dima Taji — Mai Oudah — Bashar Alhafni — Go Inoue — Fadhl Eryani — Alexander Erdmann — Nizar Habash. *CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing*. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille: European Language Resources Association, 2020, p. 7022–7032 <https://aclanthology.org/2020.lrec-1.868>.

Pakhale 2023 — Kalyani Pakhale. *Large Language Models and Information Retrieval*. «SSRN Scholarly Paper», 18 dicembre 2023. <https://doi.org/10.2139/ssrn.4636121>.

Qaroush et al. 2020 — Aziz Qaroush — Abdalkarim Awad — Mohammad Modallal — Malik Ziq. *Segmentation-based, omnifont printed Arabic character recognition without font identification*. «Journal of King Saud University - Computer and Information Sciences», 34 (2020) n. 6, p. 3025-3039. <https://doi.org/10.1016/j.jksuci.2020.10.001>.

Ramaḍān 2014 — Ibrāhīm ʿAbd al-karīm Mūsà Ramaḍān. *ʿalāmāt al-waqf fi-l-maṣāḥif al-maṭbūʿa. In: nadwa ṭibāʿa al-qur'ān al-karīm wa našrihi. al-Madīna al-munawwara*, 2014, p. 1603–1668 <https://ebook.univeyes.com/86194/pdf- علامات الوقف في المصاحف المطبوعة >.

Rice — Kanai — Nartker 1993 — Stephen Rice — J. Kanai — Thomas Nartker. *An Evaluation of OCR Accuracy*. In: *Annual Research Report*. Las Vegas: University of Nevada, Information Science Research Institute, 1993, p. 9-34.

Romanov et al. 2017 — Maxim Romanov — Matthew Thomas Miller — Sarah Bowen Savant — Benjamin Kiessling. *Important New Developments in Arabographic Optical Character Recognition (OCR)*. arXiv, 28 Mar 2017. <https://doi.org/10.48550/arXiv.1703.09550>.

Saber et al. 2016 — Shimaa Saber — Ali Ahmed — Ashraf Elsisi — Mohiy M. Hadhoud. *Performance Evaluation of Arabic Optical Character Recognition Engines for Noisy Inputs*. In: *The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015, Beni Suef, Egypt)*. Cham: Springer, 2016, p. 449–459. <https://doi.org/10.1007/978-3-319-26690-9_40>.

Saber — Ahmed — Hadhoud 2014 — Shimaa Saber — Ali Ahmed — Mohy Hadhoud. *Robust metrics for evaluating arabic OCR systems. In: International Image Processing, Applications and Systems Conference (IPAS)*. Piscataway (NJ): IEEE, 2014, p. 1–6. <https://doi.org/10.1109/IPAS.2014.7043272>.

Sadrzadeh — Muskens 2018 — Mehrnoosh Sadrzadeh — Reinhard Muskens. *Static and Dynamic Vector Semantics for Lambda Calculus Models of Natural Language*. «Journal of Language Modelling», 6 (2018), n. 2, p. 319–351.

Schimmel — Rivolta 1992 — Annemarie Schimmel — Barbara Rivolta. *Islamic Calligraphy*. Leiden: Brill Archive, 1992.

Sībawayhi 1989 — ʻAbū bašar ʿAmrū bin ʿŪṯmān Sībawayhi. Kitāb Sībawayhi. al-Qāhira: Maktabat al-ẖānğī, 1988.

Smith — Cordell 2018 — David Smith — Ryan Cordell. *A Research Agenda for Historical and Multilingual Optical Character Recognition*. DRS, 2018. <https://repository.library.northeastern.edu/files/neu:f1881m035>.

Sublet 1999 — Jacqueline Sublet. *Ḥiṣn al-ism : qirā'āt fī al-asmā' al-'arabiyya*, trad. di S. M. Barakat. Dimašq: Institute François de Damas, 1999.

Šukrī 2014 — ʻAḥmad Ḫālid Yūsif Šukrī. *alāmāt al-ẓabṭ fi-l-maṣāḥif: bayn al-wāqiʿ wa-l-ma'mūl. In: nadwa ṭibāʿa al-qur'ān al-karīm wa našrihi. al-Madīna al-munawwara*, 2014, p. 1483–1552. <https://ebook.univeyes.com/85064/pdf-الضبط في المصاحف بين الواقع والمأمول علامات>.

Taji et al. 2018 — Dima Taji — Salam Khalifa — Ossama Obeid — Fadhl Eryani — Nizar Habash. *An Arabic Morphological Analyzer and Generator with Copious Features*. In: *Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology*. Brussels: Association for Computational Linguistics, 2018, p. 140–150. <https://doi.org/10.18653/v1/W18-5816>.

Tang et al. 2024 — Jingqun Tang — Chunhui Lin — Zhen Zhao — Shu Wei — Binghong Wu — Qi Liu — Hao Feng et al. 2024. *TextSquare: Scaling up Text-Centric Visual Instruction Tuning*. arXiv, 9 Apr 2024. <https://doi.org/10.48550/arXiv.2404.12803>.

Taouka — Coltheart 2004 — Miriam Taouka — Max Coltheart. *The Cognitive Processes Involved in Learning to Read in Arabic*. «Reading and Writing», 17 (2004), n. 1, p. 27–57. <https://doi.org/10.1023/B:READ.0000013831.91795.ec>.

ISO 233-2 1993 — Technical Committee ISO/TC 46. *International Standard ISO 233-2: Information and Documentation — Transliteration of Arabic Characters into Latin Characters. Part 3: Arabic Language — Simplified Transliteration*. Geneva: ISO, 1993. <https://cdn.standards.iteh.ai/samples/78514/749948ae77474f7fa057a6b278281dbb/ISO-233-3-2023.pdf>.

ISO 233-3 2023 — Technical Committee ISO/TC 46. *International Standard ISO 233-3 : Information and Documentation — Transliteration of Arabic Characters into Latin Characters. Part 3: Persian Language — Transliteration*. Geneva: ISO, 2023. <https://cdn.standards.iteh.ai/samples/78514/749948ae77474f7fa057a6b278281dbb/ISO-233-3-2023.pdf>.

Wagaa — Kallel — Mellouli 2022 — Nesrine Wagaa — Hichem Kallel — Nédra Mellouli. *Improved Arabic Alphabet Characters Classification Using Convolutional Neural Networks (CNN)*. «Computational Intelligence and Neuroscience», 11 January 2022. <https://doi.org/10.1155/2022/9965426>.

Xu et al. 2023 — Derong Xu — Wei Chen — Wenjun Peng — Chao Zhang — Tong Xu — Xiangyu Zhao — Xian Wu — Yefeng Zheng — Enhong Chen. *Large Language Models for Generative Information Extraction: A Survey*. arXiv, 9 dicembre 2023. <https://doi.org/10.48550/arXiv.2312.17617>.

Zalaṭ bin Raf'at 2006 — Muḥammad bin Zalaṭ bin Raf'at. *'aḥkām al-tağwīd wa-l-tilāwa.* al-Qāhira: Mu'assasa qurṭuba, 2006. <https://www.noor-book.com/كتاب 0015 كتاب احكام التجويد والتلاوه محمد رافت زلط-pdf>.

Zoizou — Zarghili — Chaker 2020 — Abdelhay Zoizou — Arsalane Zarghili — Ilham Chaker. *A New Hybrid Method for Arabic Multi-Font Text Segmentation, and a Reference Corpus Construction*. «Journal of King Saud University - Computer and Information Sciences», 32 (2020), n. 5, p. 576–582. <https://doi.org/10.1016/j.jksuci.2018.07.003>.

Invented archives. Un problema teoretico per la ricerca storica o per l'archivistica?

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00151

Salvatore Spina
Università degli Studi di Catania. Dipartimento di Scienze Umanistiche

Il saggio guarda al fenomeno degli "archivi inventati" e alle sue implicazioni nella scienza archivistica e nella ricerca storica nell'era digitale. Questi archivi, per lo più tematici, emergono come risposte alla digitalizzazione e al bisogno di accesso aperto, sfidando i principi classici dell'archivistica, come il rapporto tra soggetto produttore e documento, l'accumulazione e la conservazione documentaria. Inoltre, il saggio cerca di individuare le origini teoretiche degli archivi inventati, analizzando come queste collezioni documentarie traggano ispirazione dalle metodologie di histoire sérielle *sviluppate negli anni Cinquanta del Novecento, dove la serialità e l'analisi quantitativa dei dati storici divennero centrali nella costruzione di una nuova metodologia storiografica. Con la diffusione delle tecnologie digitali, queste metodologie si sono potenziate, portando alla creazione di complessi documentari che non seguono le gerarchie archivistiche tradizionali, ma guardano alla flessibilità e alla riorganizzazione dei dati.*

1. Introduzione

L'invenzione di internet e l'implementazione del World Wide Web hanno dato all'umanità l'opportunità di estendere e rendere capillare il sistema linguistico e i meccanismi di interscambio, dove l'informatica, nata come estensione della logica e della matematica, diventa un codice di comunicazione. E se è vero che tale codice è stato sviluppato per impartire ai computer i comandi necessari affinché svolgano dei lavori di calcolo, o per attivare processi meccanici e di routine, è ancor più vero che ha spinto alla creazione della Nicchia Ecologica Digitale[1] e di una nuova *Heimat* comunicativa[2], una *interplay dimension*[3], dove la *computer science* costituisce il *comun denominatore linguistico* tra le scienze.

[1] Salvatore Spina, *Homo-Loggatus. The anthropological condition of historians in the digital world*, «Journal of Mathematical Techniques and Computational Mathematics», 10 (2023) n. 2, p. 431–437; Id, *"Homo-loggatus": uomo e macchine nella nicchia ecologica digitale*. «Informatica Umanistica e Cultura Digitale: il blog dell'AIUCD», 19 luglio 2023, <https://infouma.hypotheses.org/2224>.

[2] Giovanni Ciofalo — Silvia Leonzi, *Homo communicans. Una specie di/in evoluzione*, Roma: Armando Editore, 2013.

[3] Salvatore Spina, *Homo-Loggatus. The anthropological condition of historians in the digital world*. cit.; Id, *Digitality as a longue durée historical phenomenon*, «Umanistica Digitale», 18 (2024), p. 1–25.

MODUS

Ma il fervore e l'entusiasmo seguiti all'avvento delle tecnologie digitali, soprattutto in ambito umanistico, di fatto, hanno creato un distacco dalla reale comprensione dei processi e dei linguaggi che sottostanno all'informatica. La nascita del web non è stata segnata dalla prospettiva della "processabilità dei dati", ma solamente dalla possibilità di rapido accesso e dalla diffusione (capillare) di materiali e dati, che, nella sostanzialità, erano privi di una struttura, di un mezzo, che consentisse agli studiosi di operare analisi informatizzate. Ciò che sembra sia sfuggito è il fatto che grazie alle ICT è stato possibile portare l'attenzione sulle "parole" e intenderle come "numeri"; e, di conseguenza, guardare a tutte quelle possibilità che si potevano aprire con l'applicazione di tali tecnologie in settori della conoscenza dove la computazione, in chiave matematica, non è mai stata considerata quale colonna portante.

Ciò che da subito fu chiaro, è che il WWW era un sistema di documenti collegati tra loro attraverso dei nessi ipertestuali e milioni di URL. Tale aspetto, nell'evidenza della nostra quotidianità e della importante letteratura di settore[4], non richiede, in questa sede, ulteriori approfondimenti e definizioni di sorta.

Se è vero che l'infrastruttura Internet garantisce l'interconnessione, è ancor più vero che i computer nascono per processare dati, e l'informatica per consentire all'operatore umano di comunicare con queste tecnologie.

Questi tre postulati (interconnessione, processabilità, linguaggio logico) - se così possiamo intenderli - saranno alla base del nostro ragionamento per comprendere la nascita degli *invented archives*, ossia quegli insiemi, basati sull'infrastruttura internet e sul web, all'interno dei quali i documenti sono correlati tra loro sulla base di un rapporto prettamente tematico, e strutturati per rispondere alle direttive della *computer science*: interconnessione e computazione.

Nel momento in cui un testo viene messo online, esso assume una struttura non lineare. Il lettore non è guidato dallo scrittore in un percorso, che, come nella dimensione tradizionalmente tipografica, si muove rigo per rigo, dal primo verso l'ultimo, dall'introduzione iniziale fino all'ultimo capitolo e agli indici, i *cybertesti*[5] hanno un'organizzazione meccanica che mette il *web user* all'interno di una condizione di dinamicità in cui le sue attività neurali vengono sottoposte a stimoli diversi e variegati, pur derivanti dallo stesso documento.

Il principio dell'interconnessione diventa, quindi, fondante della struttura dei siti web, dei documenti in essi contenuti e di tutti quei complessi testuali che mettono in relazione le proprie parti, verso l'interno e verso l'esterno.

Non c'è, nella rete, un interesse diverso. Tutto è, e deve essere, *linkato*.

Da un altro lato, per ottemperare al principio della processabilità, tutto quello che ha

[4] Consiglio la lettura di Alberto Clerici — Maurizio De Pra, *Informatica e web*, Milano: EGEA, 2012; Teofilo Gonzalez et al., *Computing Handbook, Third Edition: Computer Science and Software Engineering*, Boca Raton-London-New York: CRC Press, 2014; Edwin D. Reilly et al., *Encyclopedia of Computer Science*, London: Nature Publishing Group, 2000; Zhiwei Xu – Jialin Zhang, *Computational Thinking: A Perspective on Computer Science*, Singapore: Springer Nature, 2022.

[5] Espen J. J. Aarseth, *Cybertext: Perspectives on Ergodic Literature*, Baltimore: Johns Hopkins University Press, 1997.

MODUS

segnato la nascita di paradigmi e statuti disciplinari viene rimesso in discussione per rispondere alla struttura logica su cui si fonda la *computer science*.
Un esempio determinante è dato dal concetto di *documento*, la cui definizione deve assumere una struttura logico-computazionale.
La tradizionale sua enunciazione, utilizzata dalla Diplomatica generale, è quella di «testimonianza scritta di un fatto di natura giuridica, compilata coll'osservanza di certe determinate forme, le quali sono destinate a procurarle fede e darle forza di prova» – anche se gli archivisti l'hanno utilizzata in un'accezione più generale e onnicomprensiva, in grado di includere fattispecie non necessariamente assimilabili ai soli documenti formali[6].
Tale enunciato pone grossi problemi quando si sposta il punto di vista sul piano informatico. Utilizzare, infatti, un termine che denoti la *funzione* del documento, ossia quella di "testimonianza", "prova", allontana il ragionamento dal soggetto dell'enunciato – il documento –, impedendo che si costruisca una definizione sulle *qualità* (in senso aristotelico) di quest'ultimo.
Una definizione più appropriata sarebbe, invece:

> «una sequenza di proposizioni in forma scritta e/o non scritta, che stabilisce le prerogative di un soggetto (persona fisica o giuridica) su un oggetto, su un bene mobile e/o immobile, su una condizione e/o uno status, sulla volontà o nolòntas, sui diritti e sui doveri conformi alle disposizioni della legge vigente all'atto del suo venire in essere»[7].

In questo caso, e probabilmente solo in questo, l'enunciato può essere sottoposto a un processo di computazione, in grado di produrre quegli effetti logici su cui si fonda la *computer science*. Cioè, in questo caso, siamo di fronte a una definizione processabile dalle macchine (intese, queste, come strumenti algoritmici); ovvero da *tools* di Intelligenza Artificiale (IA), che elaborano correttamente solamente quando l'informazione è *discreta* e *formalizzata*[8], ossia quando la parola è assimilabile a un'*astrazione di secondo ordine*, un *numero*[9].

---

[6] Maria Rosaria Guercio, *Archivistica informatica. I documenti in ambiente digitale*, Roma: Carocci, 2019, p. 21.

[7] La definizione è dell'autore.

[8] Tito Orlandi, *Informatica umanistica*, Roma: La Nuova Italia Scientifica, 1990; *Discipline umanistiche e informatica. Il problema della formalizzazione*, a cura di T. Orlandi, Roma: Accademia Nazionale dei Lincei, 1997; Tito Orlandi, *Formal Methods in the Humanities and their Teaching*, presentato al *The future of the humanities in the digital age*, Bergen, 1998.

[9] Leo Corry, *David Hilbert and the Axiomatization of Physics (1898–1918): From Grundlagen der Geometrie to Grundlagen der Physik*, Londra: Kluwer Academic Publishers, 2004; Gottlob Frege, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry Into the Concept of Number*, Evanston: Northwestern 1980; Id, *Senso, funzione e concetto. Scritti filosofici 1891-1897*, Bari: Laterza, 2007; Id, *L'alfabeto del pensiero*, Roma: Lit Edizioni, 2021; Id, *On Sense and Reference*, Londra: Routledge, 2011; Kurt Gödel, *The Consistency of the Axiom of Choice and of the Generalized Continuum-hypothesis with the Axioms of Set Theory*, Princeton (N.J.): Princeton University Press, 1940; Id, *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, New York: Dover, 1992; David Hilbert, *Grundzüge der Theoretischen Logik*, Berlin: Springer, 1972; Id, *Mathematical Problems. Lecture delivered before the International Congress of Mathematicians at Paris in 1900*, Lancaster & New York: Macmillan Company, 2021; David Hilbert — Stephan Cohn-Vossen, *Geometry and the Imagination*, Providence (RI): AMS Chelsea publishing, 1999; Francisco Rodriguez-Consuegra, *Kurt Gödel: Unpublished Philosophical Essays*, Basel-Boston-Berlin: Birkhäuser Verlag, 1995; Joan Rosellò. *Hilbert, Göttingen and the Development of Modern Mathematics*, Newcastle upon Tyne: Cambridge Scholars Publishing, 2019; Alfred North Whitehead et al., *Principia mathematica: Vorwort und Einleitungen*, Frankfurt am Main: Suhrkamp, 1986.

MODUS

Riconoscendone la struttura di base, quindi, ossia "le proposizioni", gli algoritmi e le IA potranno riconoscere cosa sia un documento e ricavare da esso quei dati necessari all'archivistica informatica per la classificazione e la metadatazione automatica.
Nel caso della definizione tradizionale di documento, invece, un eventuale *prompt* dato alla macchina, dovrebbe contenere tutte quelle informazioni (cosa è il soggetto, cosa l'oggetto, la struttura, la funzione ecc.) e istruzioni (ad esempio, "non usare la *funzione* per definire l'*oggetto*") che creerebbero, in realtà, risultati non controllabili, ovvero – come vengono definiti, oggi, nel contesto delle IA e dei *Large language model* (LLM) –, «allucinati»[10]. Sarebbe come impartire al calcolatore l'informazione che "il treno *trena*".
Per questo motivo, occorrerebbe che in progetti come *Records in Contexts*, o *InterPARES Trust AI (2021-2026)*, si guardasse con più attenzione ai postulati che abbiamo individuato, in quanto il mondo degli archivi, siano essi fisici, digitali, *invented* e partecipativi, vive una "sconnessione" dalle prospettive reali della Nicchia Ecologica Digitale, ossia da quello spazio caratterizzato dall'esistenza di più enti cognitivi, *agents* connessi secondo la relazione *person2persons2machines*, che non sono solamente umani, e la cui semantica è anche quella delle IA, di cui ChatGPT e la sua azienda OpenAI rappresentano, ad oggi, l'esperienza più importante.
Alla luce di tali considerazioni, nel nostro caso in esame, abbiamo la necessità di rivedere il concetto di *invented archive*, il quale, ad oggi, viene utilizzato in maniera non corretta. Ciò che accade, da un lato, perché l'aspetto della "computazione" è totalmente inesistente in tali insiemi documentari, che rappresentano, per lo più, dei complessi fotografici (quando presente l'immagine del documento), ovvero dei *dataset*, quando troviamo metadati e tags; e, da un altro lato, perché occorre uscire fuori dalla logica dell'*accumulo* e dell'*accesso libero* quali prospettive uniche del processo di digitalizzazione del patrimonio archivistico e culturale.
Esistono, infatti, svariate aggregazioni definite *invented archives*, le quali vengono costruite con l'intento di dinamizzarne i contenuti, ma che, in realtà, seppur pensate e pretese a tale scopo, non riescono a sfuggire alla tradizionale logica dell'indicizzazione, ossia di consentire agli utenti di individuare i documenti all'interno del sito, finendo col realizzare complessi non processabili. Basti pensare ai portali tematici del Sistema Archivistico Nazionale (SAN), che consentono una consultazione dal gusto totalmente retrò (analogico, di *close reading*), in cui il concetto di "calcolo" è fuori da ogni portata. Oppure al "The September 11 digital archive"[11], più volte preso quale esempio di archivio inventato da Federico Valacchi[12].
Si sta assistendo, cioè, a un processo di digitalizzazione che, anche quando accompagnato da fasi di trascrizione integrale dei testi fotografati, risponde al bisogno – non informatico! – di creare dei metadati *minimal* che consentano ai visitatori del sito inter-

[10] Takeshi Kojima et al., *Large Language Models are Zero-Shot Reasoners*, arXiv, 2023, <https://arxiv.org/abs/2205.11916>.
[11] https://911digitalarchive.org.
[12] Federico Valacchi, *Se l'archivio è artificiale: verso uno ius archivi partecipativo?*, «AIDAinformazioni», (2023) 1–2, p. 153–170.(2023

net di individuare il documento che serve loro. E poco importa se trascritto in digitale, metadatato nei contenuti, taggato, organizzato per essere "legato" ai *linked open data*; non siamo di fronte al processo di digitalizzazione che guarda al documento, nella rete, come cybertesto, così come era emerso dalle idee di Aarseth e del suo *Cybertext: perspectives on ergodic literature* – una riflessione che già nel 1965 aveva spinto Ted Nelson verso il concetto di *hypertext*.

Ciò deve portarci alla considerazione che i paradigmi su cui si fondano la *digital history*[13] e l'archivistica digitale – che non si intende, qui, quale sinonimo di "archivistica informatica", la quale guarda al problema della gestione documentale come frutto dell'attività di un soggetto giuridico – sono tutte da definire, sia dal punto di vista statutario che metodologico. E non bastano, ancora, gli accorati incoraggiamenti e inviti a guardare alla *digitalità* quale cambiamento che fa della ricerca, sia storica che archivistica, una dimensione di "collaborazione" con gli strumenti informatici e le IA – e prova ne è la difficile definizione del concetto di "archivio inventato" e la sua posizione all'interno dei paradigmi delle due discipline.

2. L'archiviazione, gli archivi e il web

Il web, scrive Nelson,

> «accepts large and growing bodies of text and commentary, listed in such complex forms as the user might stipulate. No hierarchical file relations were to be built in; the system would hold any shape imposed on it. It would file texts in any form and arrangement desired […]. Besides the file entries themselves, it would hold commentaries and explanations connected with them»[14].

Questa affermazione fa della *computer science* una grande opportunità per tutte quelle discipline che studiano il testo, come la filologia e la linguistica (computazionale), mentre archivistica e storia sembrerebbero lasciare uno spazio limitato e controllabile alle ICT. In realtà, testi e documenti, nei due settori, hanno una loro specifica valenza ma del tutto diversa nelle loro prospettive metodologiche rispetto alle scienze fisiche e naturali – o, per lo meno, questo è quello che si è cercato di sottolineare quando si è giunti all'affermazione che, essendo l'uomo l'oggetto della storia e delle scienze umanistiche in genere, gli eventi non possono essere catalogati e analizzati alla stregua di un fenomeno naturale, quindi processabile come un dato matematico.

Dal suo canto, l'archivistica ha iniziato, sin dagli albori di internet, ad acquisire uno status di presa di coscienza verso questa nuova tecnologia, che si tradusse in diverse tavole rotonde in cui numerosi studiosi presero posizione nei confronti della emergente *computer science* – tra le più importanti sicuramente il convegno del 17-19 giugno 1985

[13] Salvatore Spina, *Digital History: metodologie informatiche per la ricerca storica*, Napoli: Edizioni Scientifiche Italiane, 2022.

[14] Ted H. Nelson, *Complex information processing: a file structure for the complex, the changing and the indeterminate*, in: *ACM '65: Proceedings of the 1965 20th national conference*, New York: Association for Computing Machinery, 1965, p. 84-100: 88.

MODUS

a Torino, dal titolo "Informatica e archivi"[15]. Ma, seppur visti come opportunità, il web e i postulati su cui si fondano i concetti di *cybertesto*, *electracy* e, ancor di più, quello di *digitality*, non furono acquisiti come un vantaggio per lo statuto disciplinare dell'archivistica, anche se oggi ci si muove in maniera diversa, come dimostra il progetto InterPares Trust AI[16] il quale - anche se i sistemi di database e network tra i dati mostravano il loro perché - si sarebbe sempre fondato sui principi di conservazione e organizzazione della produzione documentaria di un soggetto, sia esso fisico o giuridico (il *soggetto produttore*, che resta chiave di volta necessaria per il venire in essere di un archivio, per garantirne l'identificazione e la sua corretta fruizione).

L'archivistica informatica, ad esempio, non ha mai ragionato in termini di computazione, perdendo la maggior parte delle opportunità che sarebbero derivate da questa prospettiva; ma, nel pieno dei dettami del suo statuto, essa guarda ai problemi della conservazione dei documenti informatici, ai requisiti funzionali per la gestione informatica di tali *corpora*, allo sviluppo di software per la classificazione e la fascicolazione, ai formati e al problema dell'autenticità. Nessuna computazione.

Ciò detto, se è vero che l'archivistica non ha mai avuto la necessità di considerare il documento (digitale) come file di testo processabile e, quindi, di considerarlo sotto la luce dell'interconnessione e della logica dei linguaggi informatici, e se è vero che gli elementi fondanti di un archivio sono il suo soggetto produttore e il vincolo archivistico, è ancor più vero che il discorso sugli *invented archives* deve partire da altre premesse concettuali.

Quando ci troviamo di fronte a un archivio inventato, dobbiamo partire dal presupposto che non abbiamo un soggetto produttore e che il *vincolo* cede il passo al *tema*; ossia ci troviamo di fronte a un atto creativo con cui un soggetto trae da archivi diversi (e, quindi, da diversi soggetti produttori) la documentazione necessaria a risolvere una questione storica, per legarla insieme senza un vincolo archivistico in senso stretto, ma allo scopo di farla diventare un complesso di fonti storiche, su cui si ergerà il ragionamento degli studiosi e la loro interpretazione storiografica.

Occorre, però, non cadere nell'errore di considerare archivistica, internet e *invented archives* dei mondi privi di un legame: gli archivi tradizionali (sia fisici che digitali) sono la base fondante degli *invented*. Come afferma Rosenzweig, gli archivi inventati sono «sites devoted to collecting and making available documents that are scattered in various "real" archives»[17], ossia questi ultimi (i *real archives*) servono alla creazione della sedimentazione degli *invented*, ma sono altro rispetto ad essi. Li anticipano, magari guardandoli come *effetti* ulteriori propri.

[15] Ministero per i Beni Culturali e Ambientali. Ufficio Centrale per i Beni Archivistici — Archivio di Stato di Torino et al., *Informatica e Archivi. Atti del Convegno. Torino 17-19 giugno 1985*, Roma: Istituto Poligrafico e Zecca dello Stato ; Libreria dello Stato, 1986.

[16] https://interparestrustai.org/trust.

[17] Roy Rosenzweig, *Brave New World or Blind Alley? American History on the World Wide Web*, «Journal Of American History», 1 (1997) n. 84, p. 132–155; Id, *The Road to Xanadu: Public and Private Pathways on the History Web*, «Journal of American History», 2 (2001), n. 88, p. 548–579.

MODUS

Dal primo sito messo online[18], col quale si dava libero accesso agli utenti del web, alle idee e alle caratteristiche proprie dell'invenzione del WWW, tutto quello che è stato "uploadato" ha poco a che vedere con il concetto di archivio. Mentre gli *invented archives* si fondano proprio sul web, in quanto vedono in questa infrastruttura il "luogo" dove riunire la documentazione che viene intercettata dai loro creatori e, allo stesso tempo, il *medium* essenziale per farla conoscere.

Dal 1991, la rete vede una costante implementazione di siti web che racchiudono documenti archivistici, la cui finalità è quella di dare corpo "digitale" a un progetto storico, a una idea di ricerca o semplicemente a una passione personale. Il "Mosaic Netscape" e il "Netscape Navigator" (era già il 1994) diedero a tutti la possibilità di creare le proprie collezioni e di fare della rete una struttura informativa globale. La "storia" era dappertutto e, sorprendentemente, le persone si precipitarono per mettere le proprie storie su internet, grazie alla creazione di siti dedicati ai loro soggetti e temi preferiti.

Non siamo di fronte ad archivi, bensì a quello che Stefano Vitali[19] definisce il «risultato dell'iniziativa di molteplici soggetti, animati da varie ed eterogenee motivazioni». Soggetti che non (cor)rispondono alle solide caratteristiche del "produttore", quanto invece alla figura del *creator*, e che al fine di soddisfare i democratici principi della "diffusione larga" aggregano e raccolgono tutta la variegata documentazione che intercettano negli archivi per legare insieme le varie carte e porle alla base di filoni di ricerca, siano essi di ampia visione oppure a supporto di un evento specifico, limitato nello spazio e nel tempo.

Certamente, c'è da sottolineare che non è solo l'entusiasmo di cultori a far sì che emergano questi *database online* – termine, a mio avviso, più indicato – ma anche la volontà di ricercatori professionisti, i quali mostrano la chiara finalità di creare un complesso documentario in grado di guidare, successivamente, la ricerca storica in quell'ambito determinato. Numerose aggregazioni vengono realizzate tra il 1995 e il 2000 e crescono continuamente, con chiari scopi scientifici, ma in assenza di un vincolo archivistico scientificamente inteso. In altre parole, risulta ben definito il criterio che determina la scelta di farvi accedere dei documenti piuttosto che altri, ma non troviamo, tra queste collezioni, alcun archivio tradizionale, e guardarle come tali pare presenti più rischi che vantaggi. Ciò è ancor più vero quando si pensa agli *invented archives* e all'infrastruttura informatica su cui si basano, ossia alle "relazioni semantiche" – espressione che non si presenta, già di per sé, di facile analisi. Parlare, infatti, di semantica vuol dire addentrarsi nelle complesse maglie della comunicazione umana: il significato, i significati.

Oggetto della ricerca umanistica non è una semplice informazione, ma la produzione visibile e invisibile della mente umana che non si struttura in regole e leggi, quanto piuttosto in segni e glifi, i quali, però, sono parte della dimensione della conoscenza; "denotatori"[20], ossia, come asserisce Sebeok, «sign[s] without either similarity or conti-

[18] http://info.cern.ch/hypertext/WWW/TheProject.html.

[19] Stefano Vitali, *Passato digitale: le fonti dello storico nell'era del computer*, Milano: Bruno Mondadori, 2004, p. 116.

[20] Ernst Cassirer, *The philosophy of symbolic forms*, London and New York: Routledge, 2021.

MODUS

guity, but only with a conventional link between [their] signifier and [their] denotata»[21], quindi, un'espressione convenzionale in cui il denotato è qualcosa di indipendente. Un assunto, questo, che mette in evidenza una prospettiva d'indagine profonda sul concetto stesso di testo, che parte dal presupposto che ogni scrittore – produttore di documenti – attribuisce a un segno, sulla base di regole che lui stesso sceglie, un significato. Ed è tale relazione che la *computer science* deve ambire a ricostruire per poter tradurre in linguaggio informatico la corretta informazione e, qualora possibile, l'intenzionalità dell'autore, ossia la sua consapevolezza che il messaggio avrà la funzione di causare una reazione[22].

In ambito informatico, quanto meno a partire dalle esperienze di Ada Lovelace e Babbage, i quali guardavano più che altro all'aspetto *discreto* dell'informazione da far processare[23], il tema dei *significati* (e, quindi, della semantica) aveva avuto poco *appeal*, e questo fino agli anni Cinquanta del Novecento e alla fatidica domanda posta da Alan Turing: «Can Machines think?»[24].

Questa domanda ha dato vita a diversi progetti di ricerca, la cui linea d'azione era quella di comprendere quali meccanismi devono crearsi e sottostare a quella che, a partire proprio da quegli anni, venne chiamata intelligenza artificiale[25]. Questo percorso ci porta – e porta gli *invented archives* – molto lontano rispetto ai paradigmi dell'archivistica, la quale, da sempre, ha basato le metodologie di conservazione e fruizione su meccanismi che non hanno mai dato particolare attenzione al significato.

Un documento non ha un "significato" perché inserito in un determinato contesto (serie, fondo ecc.) come per gli archivi inventati, ma esso si trova in un determinato "frangente" (inteso come *momento* spaziotemporale) dell'archivio, perché lì sta la sua ragione giuridica, e perché lì il soggetto produttore ha deciso di collocarlo. Non vi è una semantica, la quale potrebbe, probabilmente, emergere sotto la spinta di analisi e procedimenti intellettuali.

L'organizzazione di un archivio è di tipo gerarchico e non relazionale. È una struttura a "grafo", ad "albero"; mentre il web e gli *invented archives* – seppur detentori di un termine (*archive*) che da un punto di vista prettamente archivistico non denota il loro reale status – chiedono altro, come abbiamo avuto modo di analizzare.

[21] Thomas A. Sebeok, *Encyclopedic Dictionary of Semiotics*, Berlin-Boston: De Gruyter Mouton, 1986, p. 108.

[22] Gianluca Braschi, *Archivi inventati*, «Medium», 19 luglio 2024, <https://braschi.cc/archivi-inventati-9d67aba2e83a>.

[23] S. Spina, *Digitality as a longue durée historical phenomenon*. cit.

[24] Alan Mathison Turing, *Computing machinery and intelligence*, «Mind», 236 (1950), n. LIX, p. 433–460.

[25] John McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955*, <https://upload.wikimedia.org/wikipedia/commons/f/f4/A_Proposal_for_the_Dartmouth_Summer_Research_Project_on_Artificial_Intelligence%2C_by_John_McCarthy_et_al%2C_1955.pdf>.

MODUS

3. Internet, archivi*, invented, web archiving*

Il 6 agosto 1991, veniva lanciato il primo sito web della storia dal fisico britannico Tim Berners-Lee – noto come l'inventore del World Wide Web. Nel decennio successivo, i siti hanno uno sviluppo esponenziale e il numero di database e complessi documentari online aumenta sempre più, fino ai nostri giorni. Occorre dunque, per ciò che ci riguarda, un *distinguo* tra le varie opportunità di consultazione che si aprono di fronte ai *web-users*, i quali possono accedere a:

1) siti ufficiali di archivi fisici e biblioteche;
2) *invented archives*;
3) siti di *web archiving*.

Nel primo caso, ci troviamo di fronte a spazi in cui gli utenti hanno la possibilità di accedere a tutte le informazioni relative alle attività dell'ente conservatore e ai fondi che esso possiede e conserva. Nell'ultimo decennio, a tali finalità si aggiunge quella della consultazione delle unità archivistiche, che scaturisce dalla riflessione degli archivisti, la quale mette al centro il "punto di vista esterno"[26] rispetto a quello degli archivi; non più, cioè, quello che l'archivio rappresentava per sé stesso (e per gli archivisti), bensì quello che esso avrebbe rappresentato per il mondo che, da quel momento, lo avrebbe "preteso" attraverso siti e repository online. Non più il sito internet come pagina ufficiale, ma come portale di accesso alla documentazione riprodotta in formato digitale e/o eventuali database che cercano di riprodurre la struttura ad albero dei fondi, talvolta cercando di garantire anche un accesso analitico grazie a dei metadati generali. Non *archivi inventati*, quindi, ma siti che rispondono, da un lato, al sistema di pubblicità degli enti conservatori e, da un altro lato, alla consultazione. E anche nel caso di una messa online di fondi archivistici o parti di essi, il concetto di *invented* non va a denotare la sostanzialità di questi progetti di digitalizzazione, ossia la creazione di versioni consultabili che non hanno, quindi, alcun nesso semantico tra i dati se non quello tra le carte, il fondo di cui fanno parte, e il vincolo che le lega al soggetto produttore – unica fonte in grado di garantire, in maniera inoppugnabile, l'attività che ha dato vita a quel determinato documento.

Avendo già chiarito ciò che concerne il secondo punto (gli *invented archives*), il nostro ragionamento può spostarsi verso il terzo elemento: il *web archiving*.

Berners-Lee aveva pensato – e realizzato – una rete di informazioni che permettesse agli utenti di accedere a contenuti (le pagine dei siti web, le immagini, i video, i vari documenti ufficiali che si trovano nei siti istituzionali dei governi, dei ministeri ecc.) grazie a un URL (Uniform Resource Locator) utile ad intercettare l'informazione.

Ma il problema che ne deriva è quello della destinazione finale di tali contenuti, dell'obsolescenza dei formati e della rivoluzione che questa tecnologia comporta nei vari ambiti della conoscenza. Scaturiscono domande che chiamano in causa gli umanisti, gli archivisti, gli storici: questo flusso di documenti ha la necessità di confluire in un archivio?

[26] Ministero per i Beni Culturali e Ambientali. Ufficio Centrale per i Beni Archivistici — Archivio di Stato di Torino et al., *Informatica e Archivi. Atti del Convegno. Torino 17-19 giugno 1985*, cit., p. 15.

Quale sarà la struttura di tale archivio? Come si rapporteranno gli storici e gli archivisti a tali contenuti?

Anche se molte di queste domande sembrano aver trovato, nel tempo, una risposta, trent'anni sono un tempo troppo breve per far sì che uno statuto millenario come quello dell'archivistica possa rapidamente porre delle concettualizzazioni definitive e chiare al problema della digitalizzazione e del trattamento/conservazione dei documenti nella rete.

Fin dai suoi albori, il web ha sentito la necessità di creare uno spazio dove "archiviarsi". Tutto quello che viene concepito come informazione, fin dalla nascita dell'umanità, entra nel corso del tempo che scorre e, logicamente, nel passato. Abbiamo l'innata necessità di archiviare tutto quello che produciamo, in termini di scrittura, concetti ed enunciati, qualunque sia il supporto che utilizziamo. Il web rappresenta un nuovo supporto, e per questo motivo, nel 1996, qualche anno dopo l'upload del primo sito internet, Brewster Kahle fonda Internet Archive, una macchina (archivistica) virtuale che mira a preservare i siti web creati e messi online, che rischiano di sparire dalla rete se gli spazi sui server dei provider non vengono più finanziati.

Siamo di fronte, dunque, a un'organizzazione non-profit (il *soggetto collettore*) costituita da varie istituzioni governative e non, che ha dato vita a una libreria con sede a San Francisco in cui convergono «millions of books, movies, software, music, websites, and more, all free of charge»[27].

Il sistema si basa su una struttura *entry point* che consente ai *web users* di individuare un contenuto e di ricostruire, nel caso di un sito web, la sua evoluzione in termini di aggiornamenti dell'interfaccia e della sua composizione grafica e funzionale originaria. A questo si aggiunga la possibilità di poter leggere ed effettuare il download di quei documenti che, dalla messa offline del sito ospitante, non sono più disponibili. Internet Archive, infatti, attraverso la sua Wayback Machine consente di intercettare nella collezione un PDF o qualunque altro file e memorizzarlo in remoto (download).

Rispetto alla sua fondazione, nel corso del tempo la piattaforma ha superato i suoi limiti geografici, estendendo l'archiviazione a tutto quello che si trovava sul web attraverso il processo di *crawling*[28].

Esiste, dunque, un soggetto conservatore? I vari archivi, una volta dentro la Wayback Machine, saranno considerati alla stregua di fondi, o ancora meglio, alla stregua di inventari? Come si rapporteranno gli storici con Internet archive?

Domande non semplici, alle quali la letteratura non sembra aver trovato risposta.

Possiamo ipotizzare, però, che ogni sito memorizzato non debba essere considerato un documento, ma un fondo specifico, preservato dall'organizzazione di Brewster

[27] https://archive.org/.

[28] L'archiviazione avviene attraverso l'applicazione di software *crawler* specializzati, i quali "scandagliano" il web, per copiare e salvare informazioni. Nell'ambito della Web Archive Collection, i siti archiviati e le informazioni che essi rappresentano sono resi accessibili online. Questi siti e le informazioni raccolte sono anche visibili, leggibili e consultabili, proprio come quando erano direttamente sul web, ma vengono conservati sotto forma di istantanee web delle informazioni in un certo momento. Dentro l'archivio è possibile trovare siti di biblioteche nazionali, archivi nazionali, organizzazioni di informazione, gruppi tecnologici e varie altre organizzazioni. Queste parti, inoltre, sono tutte coinvolte nella conservazione accurata del materiale online più significativo.

MODUS

Kahle, che può fungere da ente conservatore, e da poter aprire e consultare nel tentativo di individuare, al suo interno, un determinato documento (pensiamo ai file PDF delle determine dei sindaci degli enti comunali, a cui si potrebbe accedere attraverso la pagina dell'albo pretorio online, salvata in Internet Archive, come mostrato nella figura che segue).

Figura 1. *Pagina dell'Albo pretorio online del Comune di Riposto (provincia di Catania, Italia*

Certamente, gli interrogativi sull'"archivio degli archivi (digitali)" restano tanti: quali sono i criteri di classificazione adottati dall'organizzazione, ben sapendo che l'ordinamento costituisce una funzione centrale per la corretta formazione e gestione di un archivio, soprattutto se prodotto in forma digitale? Il motore di ricerca di Internet Archive e lo strumento Wayback Machine costituiscono un *thesaurus*, ovvero un inventario *tout court*?

A tutto questo si aggiunga che il sito non fornisce un motore di ricerca per i contenuti dei siti e non fornisce informazioni bibliografiche esaurienti, o brevi, sui documenti che contiene. L'archivio, inoltre, non divulga dettagli sui siti che conserva, la quantità e la natura dei loro nomi di dominio e il paese a cui appartengono. Ma la criticità più importante è che, poiché l'archivio non fornisce alcuna categorizzazione o classificazione oggettiva dei siti che preserva, l'utente deve essere a conoscenza dell'URL per poter individuare il documento "archiviato" dentro l'enorme *heritage*. Ciò non gioca a favore dell'accessibilità rapida, ma, soprattutto, dei criteri generali e standardizzati che sottostanno al concetto di collocazione (ad esempio stanza, scaffale, faldone, fascicolo ecc.), stabiliti dalla dottrina archivistica. Come afferma Stefano Allegrezza, «i documenti che riguardano uno stesso affare formano un fascicolo e, nel mondo analogico, vengono collocati all'interno di una camicia [sulla quale] vanno riportati i dati identificativi del fascicolo», e più fascicoli «sono generalmente riuniti in un contenitore (cartelle, buste, mazzi, faldoni)». E il mondo digitale non si discosta da tale metodologia, se non nella terminologia utilizzata, così come, ancora, descrive Allegrezza:

> «Passando al mondo digitale, i documenti cartacei vengono 'rimpiazzati' da documenti informatici, che, in ultima analisi, sono file — e quindi sequenze di *bit* — e come tali necessitano, anche solamente per 'esistere', di un supporto su cui essere memorizzati. I fascicoli diventano 'cartelle' o altre forme di aggregazione documentale nel *file system* in uso, mentre il concetto di unità di condizionamento, stante la diversa 'fisicità' dei materiali digitali, perde molta della sua importanza. Gli scaffali scompaiono per lasciare il posto ai *rack*, strutture metalliche su cui vengono collocati i dispositivi di storage — come gli *hard disk* o i *solid state disk* (SSD) che stanno progressivamente soppiantando i primi — e l'infrastruttura necessaria per il loro funzionamento e che, nel loro complesso, consentono di memorizzare i documenti informatici. I locali degli archivi tradizionali vengono rimpiazzati dai datacenter che hanno anch'essi la necessità di adottare tutta una serie di misure di sicurezza sia logiche che fisiche per garantire la conservazione degli archivi digitali in essi ospitati al pari dei loro equivalenti analogici»[29].

Internet Archive non risponde a tale logica. Abbiamo un ente conservatore, ma non abbiamo un sistema di classificazione e un inventario che aiutino l'utente a individuare l'unità archivistica. E su questo concetto, gioca negativamente anche il fatto che, nel web, l'URL sembra essere l'unica nuova concettualizzazione di collocazione. Si aggiunga che la Wayback Machine consente di individuare il sito internet – quello che potremmo definire "fondo" – ma non il suo contenuto. Non vi è, quindi, un inventario costruito dall'ente conservatore, e l'unico modo per scoprire il contenuto dei fondi è quello di navigare tra le pagine di un sito archiviato, senza poter utilizzare il motore di ricerca interno al sito stesso, il quale, non lavorando più sul server originario, non può accedere al dataset. L'unico modo per intercettare una "unità" è conoscerne esattamente l'URL.
Nel caso già preso in esame come esempio: il sito internet del comune di Riposto è

[29] Stefano Allegrezza, *Towards a new archival economy: the development of the discipline in the transition from analogue to digital*, «JLIS.It», 1 (2017), n. 8, p. 114–126.

MODUS

<http://comune.riposto.ct.it/>, ma dentro la Wayback Machine l'URL (la collocazione in ambiente digitale) per la sua individuazione diventa <https://web.archive.org/web/20130913175143/http://www.comune.riposto.ct.it/it/elenco-albo-pretorio.php> – che sarà, nel brevissimo futuro, alla base delle ricerche di quegli studiosi che avranno interesse a scrivere la storia di questo ente (o di altri) o di eventi ad esso collegati.

4. La rete e la ricerca storica. Gli *invented archives* come *histoire sérielle*?

Se è vero che gli *invented archives* non presentano quelle caratteristiche che ci consentono di accostarli a un archivio tradizionale (anche se costruito in uno spazio web), è ancor più vero che l'individuazione di una genealogia di tale concetto può trovarsi solamente in ambito storico, spostando la nostra analisi verso il metodo storiografico dell'*histoire sérielle*, ossia dentro quell'approccio che negli anni Cinquanta del Novecento fondava la sua metodologia sull'individuazione di determinati documenti archivistici e il loro inserimento in una *serie* definita, allo scopo di spiegare un fenomeno storico che diventava, per Furet[30], condizione necessaria per il venire in essere della storia quantitativa[31].

A servirsi dell'espressione *histoire sérielle* è stato Pierre Chaunu, per chiarire la prospettiva della sua opera *Séville et l'Atlantique, 1504-1650*[32]. Seguiranno, poi, altre sue opere che specificheranno ancora tale concetto, come *Histoire quantitative ou histoire sérielle* e *La durata, lo spazio e l'uomo nell'epoca moderna. La storia come scienza sociale*[33], la quale impegnava il lettore all'analisi dell'elemento ripetuto – che si integrava in una dimensione di omogeneità – e non al fatto individuale, con la finalità di costruire un complesso di informazioni processabili, secondo le metodologie delle matematiche sociali.

La serialità sostituisce all'inafferrabile avvenimento della storia positivistica, la ripetizione regolare dei dati selezionati, lasciando allo storico la possibilità di muoversi dentro il complesso di documenti, il quale acquisisce una struttura livellare, suddivisa in sottosistemi, «di cui egli è libero poi di proporre, o meno, le articolazioni interne»[34] al fine di rispondere al suo quesito sulla storia.

Siamo agli albori della *Digital History*, ma soprattutto di quello che, oggi, è il concetto di *invented archive*.

---

[30] François Furet, *L'histoire quantitative et la construction du fait historique*, «Annales. Histoire, Sciences Sociales», 1 (1971), n. 26, p. 63–75; Id, *Il quantitativo in Storia*, in: *Fare storia: temi e metodi della nuova storiografia*, a cura di J. Le Goff e P. Nora, Torino: Einaudi, 1981.

[31] Robert Darcy — Richard C. Rohrs, *A guide to quantitative history*, Westport, Conn.: Praeger, 1995; Charles M. Dollar — Richard J. Jensen, *Historian's guide to statistics; quantitative analysis and historical research*, Huntington, N.Y.: R.E. Krieger, 1974; François Furet, *Sur quelques problèmes posés par le développement de l'histoire quantitative*, «Social Science Information», 1 (1968), n. 7, p. 71–82; Oscar Itzcovich, *Dal mainframe al personal, il computer nella storiografia quantitativa*. in: *Storia e Computer: alla ricerca del passato con l'informatica*, a cura di S. Soldani e L. Tomassini, Milano: Mondadori, 1996, p. 30-47.

[32] Pierre Chaunu, *Séville et l'Atlantique, 1504-1650 : Structures et conjoncture de l'Atlantique espagnol et hispano-américain (1504-1650). Tome II, volume 1 : La conjoncture (1504-1592)*, Paris: SEVPEN, 1959.

[33] Pierre Chaunu, *Histoire quantitative ou histoire serielle*, Paris: Armand Colin, 1964; Id, *La durata, lo spazio e l'uomo nell'epoca moderna. La storia come scienza sociale*, Napoli: Liguori, 1983.

[34] François Furet, *Il quantitativo in Storia*. cit., p. 6.

MODUS

Gli anni Novanta del Novecento sono forieri di novità metodologiche in ambito storico. L'avvento delle tecnologie informatiche rimodula e porta la riflessione verso nuove prospettive, su cui si aprono crisi e diffidenze[35], ma anche visioni di progressione nella scienza storica. Se da un lato, infatti, Braudel sottolineava la «crisi generale delle scienze dell'uomo»[36], dall'altro accettava la necessità di ragionare sui fenomeni per inserirli in un'originale visione dei tempi: la breve, la media e la lunga durata.

In tutti questi casi, seppur non evidente, la presenza di dati costanti è necessaria non solo alla costruzione della teoria, bensì anche a quella storiografia che individua elementi "onnipresenti" nel tempo. Quest'ultimo, chiamato in causa sulle trame di ampie partizioni che non possono non basarsi sull'analisi di elementi/dati che, una volta individuati, spiegano la serie, la quantità, la lunga durata. Ed è ancora Furet ad affermare che «la storia seriale [è] una delle strade più feconde dello sviluppo della conoscenza storica»[37], non già solamente per gli apporti quanto per l'aver posto le condizioni necessarie al venire in essere di nuovi concetti che, oggi, dopo maturi dibattiti, hanno mostrato il lato "promettente" della digitalità: dataset, database, data organization, Big Data, digital archive, e così via.

Nell'intimità della sua ricerca, ogni storico organizza le sue informazioni, isola dei dati e li inserisce in un prodotto storiografico, nel quale prende corpo la sua interpretazione su eventi e soggetti del passato.

Sia essa una prova evenemenziale o un fenomeno di lunga durata, ciò che caratterizza il lavoro dello storico, nelle sue prime fasi, è la necessità di organizzare e concatenare le informazioni che trae dal *close reading*.

Questa necessità si era concretizzata, sotto la sollecitazione delle idee che Vannevar Bush fa confluire nel suo *As We May Think*[38], già a partire dagli anni Cinquanta, nell'uso di computer e software in grado di organizzare le informazioni storiche in dataset.

Il database, nel pensiero di Oscar Itzcovich, diventa lo strumento più duttile e tecnologicamente avanzato, per lo storico, per costruire complessi di informazione e analizzarli[39] superando l'approccio di *close reading*.

La matematizzazione della documentazione scritta, in tale prospettiva, è una prassi che si fa sempre più presente nella ricerca umanistica. La maggior parte delle fonti d'archivio può essere datificata[40] e inserita in *mainframe*, database e archivi inventati.

Il Novecento, però, ha mostrato ben altre possibilità. Assunto che il database abbia fornito la base concettuale degli *invented archives*, la presenza di questi ultimi si fa più capillare grazie alla infrastruttura internet e al web.

Il 20 dicembre 1990, il CERN mette a disposizione dei suoi dipendenti il progetto WWW.

---

[35] Robert Rowland, *L'Informatica e il mestiere di storico*, «Quaderni storici», 3 (1991), n. 26, p. 693–720.

[36] Fernand Braudel, *Scritti sulla storia*, Milano: Bompiani, 2003.

[37] François Furet, *Il quantitativo in Storia*, cit., p. 4.

[38] Vannevar Bush, *As we may think*, «Atlantic», 1 (1945), n. 176, p. 112-124.

[39] Oscar Itzcovichk, *Lo storico e il database*, «Quaderni storici», 70 (1989), n. 24, p. 321–325; Id, *L'uso del calcolatore in storiografia*, Milano: FrancoAngeli, 1993.

[40] Salvatore Spina, *Datificazione delle fonti storiche per la digital history delle pandemie*, «Umanistica Digitale», (2021), n. 10, p. 1-18.

MODUS

Il 6 agosto 1991, come già detto, appare nella rete il primo sito internet e il 30 aprile 1993 il CERN decide di rendere il World Wide Web *opensource*, rilasciando il codice sorgente in pubblico dominio e determinando una sostanziale rottura nei modi di produzione e di comunicazione dei risultati delle ricerche, finanche quella umanistica e storica. In Italia, il primo sito web fu messo online dal Centro di ricerca, sviluppo e studi superiori in Sardegna (CRS4)[41], oggi ancora attivo, nella primavera del 1993.

Il web ha modificato – e continua a farlo in quei settori in cui ci si trova in profondo ritardo con l'acquisizione della prospettiva informatica – la vita dei ricercatori che vengono investiti, proprio in quel periodo, dall'impossibilità di restare fuori da questo *metamedium*[42] che irrompe nella vita quotidiana di ciascuno in misura crescente, in relazione all'aumento delle possibilità di interazione, influenzando non solo i modi e i tempi della ricerca storica, ma ancor più profondamente le pratiche di disseminazione e di condivisione del suo discorso, la didattica della storia e la sua comunicazione[43].

La ricerca sul passato ha assunto tutto un altro senso metodologico, non semplice da definire: «Non esistono in sostanza prodotti già bell'e pronti, adatti all'impiego da parte degli storici, né forse esisteranno mai»[44].

È qui che trova posto e si espande il concetto di *invented archive*: tutto può essere dentro la rete, può essere veicolato e, soprattutto, arricchito. La rete è un mondo aperto, di condivisione, che porta sul *desk* degli studiosi una quantità di materiale inimmaginabile[45] e, il più delle volte, incontrollabile.

Ecco, quindi, il nodo della nostra disamina: posto che l'archivistica inizia a guardare agli *invented archives* con particolare attenzione, sembrerebbe tuttavia non aver elaborato una specifica concettualizzazione di "archivio inventato", ed è la ricerca storica, invece, a sostenere il fondamento e il venire in essere di tale concettualizzazione in maniera costante. La storia è dappertutto. Storici e studiosi di vario livello mettono online grossi complessi documentari che attirano l'attenzione di altri storici e altri studiosi, i quali necessitano di accedere a quei documenti.

Nel 2000, Dennis Trinkle e Scott Merriman danno alle stampe il libro *The History Highway 2000*[46], la cui lettura e consultazione può dare un'idea della voglia di contribuire

[41] https://www.crs4.it.

[42] Klaus Bruhn Jensen, *Metamedium*. *The International Encyclopedia of Communication Theory and Philosophy*, Chichester: Wilet-Blackwell, 2016, p. 1–4.

[43] Enrica Salvatori, *Digital (Public) History: la nuova strada di una antica disciplina*, «RiMe. Rivista dell'Istituto di Storia dell'Europa Mediterranea», (2017), p. 57–94.

[44] Robert Rowland — Renzo Derosas, *Premessa*, «Quaderni Storici», 78 (1991), n. 3, p. 687-691: 691.

[45] Syed E. Ahmed, *Perspectives on big data analysis methodologies and applications*, ed. by International Workshop on Perspectives on High-Dimensional Data Analysis, Providence (RI): American Math. Soc., 2014; Joris van Eijnatten et al., *Big Data for Global History: The Transformative Promise of Digital Humanities*, «BMGN - Low Countries Historical Review», 4 (2013), n. 128, p. 55–77; Roberto Franzosi, *A third road to the past? Historical scholarship in the age of big data*, «Historical Methods: A Journal of Quantitative and Interdisciplinary History», 4 (2017), n. 50, p. 227–244; Shawn Graham et al., *Exploring Big Historical Data. The Historian's Macroscope*, Singapore: World Scientific Publishing, 2015; Richard J. Lane, *The big humanities: digital humanities/digital laboratories*, New York: Routledge, 2017.

[46] Dennis A. Trinkle — Scott A. Merriman, *The history highway 2000, a guide to Internet resources*, Armonk (N.Y.): M.E. Sharpe, 2000.

alla costruzione di una massa critica che renda la ricerca storica un'attività sempre più dinamica e partecipativa – è questo ultimo il principio alla base della *public history*.
È la rivalsa dell'*histoire sérielle*? Probabilmente. Così come i database offline, utilizzati originariamente dagli studiosi, gli *invented archives* richiedono un processo di datificazione e metadatazione dei vari documenti in essi inseriti (record). Non basta, quindi, scegliere un tema storico e muoversi alla ricerca della documentazione archivistica (sia pubblica che privata) per costruire un archivio inventato. Il tema è il *leitmotiv* ma la funzione del database/archivio digitale inventato è quella di costruire un complesso di dati che possano, *in primis*, dialogare tra loro ed essere processabili per consentire allo studioso di acquisire informazioni sempre più analitiche e individuare modelli che possano spiegare un determinato evento, o una serie di essi, e che rappresentino l'esplicazione del tema fondante del complesso documentario.
Gli archivi inventati, dunque, la loro sempre più rapida proliferazione, oltre che lo sviluppo continuo di applicazioni e *tools* per la loro creazione, come Omeka-S[47] e Arches[48], affascinano e attraggono anche gli studiosi legati alla tradizione della metodologia qualitativa, i quali, come affermava già Furet, vedono un'efficacia superiore nella serialità e nella quantificazione[49], soprattutto quando la presenza di archivi e documentazione e l'assenza di interruzioni continue, come nel caso della storia moderna, ne valorizzano l'approccio.
Gli *invented archives* sono frutto spontaneo della metodologia della serialità al servizio degli storici intenti a "misurare" il cambiamento e i mutamenti, l'origine dei fenomeni e la loro diffusione geografica. Pensiamo, ad esempio, ai database sul fenomeno della schiavitù, uno tra tutti Slave Voyages[50], i quali mostrano la sua complessità e la capillarità.
Avendo, lo storico, la necessità di organizzare le sue carte per individuare quelle che possano assumere il ruolo di fonte per l'analisi e la descrizione del suo problema storico, in ambiente digitale, il processo di datificazione diventa molto più profondo e si traduce nella creazione di una versione dinamica di un documento, un *cybertesto*.
Dentro gli archivi inventati, la documentazione risponde alle necessità dello storico e non all'"accessibilità" archivistica. Non conta solamente la sua collocazione, ma la sua relazione con le altre fonti del database, anche se lo stesso Records in Contexts (RiC) guarda a tale prospettiva.
Al suo interno, l'archivio inventato:

> «would accept large and growing bodies of text and commentary, listed in such complex forms as the user might stipulate. No hierarchical file relations were to be built in; the system would hold any shape imposed on it. It would file texts in any form and arrangement desired - combining, at will, the functions of the card file, loose-leaf notebook, and so on. It would file under an unlimited number of categories. It would provide for filing in Bush trails. Besides the file entries themselves,

[47] https://omeka.org/s/.
[48] https://www.archesproject.org.
[49] François Furet, *Il quantitativo in Storia*, cit., p. 6.
[50] https://www.slavevoyages.org.

> it would hold commentaries and explanations connected with them. In addition to these static facilities, the system would have various provisions for change. The user must be able to change both the contents of his file and the way they are arranged. Facilities would be available for the revising and rewording of text. Moreover, changes in the arrangements of the file's component parts should be possible, including changes in sequence, labelling, indexing and comments»[51].

L'archivio inventato deve rispondere, infatti, alla dinamicità della ricerca storica, la quale modula la sua fonte sulla base del suo progetto di ricerca. La serialità, infatti, spinge lo storico a non esaurire mai le possibilità dell'interpretazione e dell'analisi delle fonti, le quali, proprio nella struttura dell'*invented archive* paragonabile a quella che Nelson definisce l'Evolutionary List File (ELF)[52], «can be used and compounded in many different ways»[53]. Ricerca storica e *invented archives* sono dimensioni strettamente legate, la cui struttura paradigmatica è stata segnata dalla serialità e dal "quantitativo" che portano lo storico a ragionare in termini diversi sulle immagini del passato. Egli si trova di fronte a un nuovo modo di organizzazione dei dati, i quali, seppur ordinati nel tentativo di farsi comun denominatore del tema che ha dato origine all'archivio inventato, rispondono a una malleabilità che fa del "fatto" un'interpretazione basata sulla coerenza tra di essi.

> «Il documento, il dato, non esistono più di per [sé] stessi, ma in rapporto alla serie che li precede e li segue; ciò che diventa obiettivo è il loro valore relativo e non il loro rapporto con un'inafferrabile sostanza "reale". Così si trova a essere spostato, al tempo stesso, anche il vecchio problema della "critica" del documento storico»[54].

L'*invented archive*, quindi, garantisce quella che Furet definisce come una critica "esterna", la quale

> «non si basa più su una credibilità fondata sul confronto del documento con testi contemporanei di natura diversa, ma sulla coerenza con un testo della stessa natura che ha una posizione diversa nella stessa serie temporale, ossia lo precede o lo segue. [Al contrario, la] critica "interna" viene ad essere tanto più semplificata, in quanto molte operazioni di "rastrellamento" dei dati possono essere inserite nella memoria del calcolatore elettronico»[55].

Ed è ricorrendo ancora alle affermazioni di Furet, che possiamo individuare una generale, decisiva e prima definizione di *invented archive*:

---

[51] Ted Nelson, *A File Structure for the Complex, the Changing, and the Indeterminate*, cit., p. 88.

[52] Ibidem, p. 89–90: «It is an evolutionary file structure constructed with lists. The system proposed here is not the only ELF possible. It is built upon a specific technique of attaching lists together which has a natural resistance to becoming confused and messy. The ELF has three elements: entries, lists and links. An entry is a discrete unit of information designated by the user. It can be a piece of text (long or short), a string of symbols, a picture or a control designation for physical objects or operations. A list is an ordered set of entries designated by the user. A given entry may be in any number of lists. A link is a connector, designated by the user, between two particular entries which are in different lists».

[53] Ibidem, p. 89.

[54] François Furet, *Il quantitativo in Storia*, cit., p. 9.

[55] Ibidem.

MODUS

«Questo incontro di una rivoluzione metodologica e di una rivoluzione tecnica, che non sono d'altra parte estranee l'una all'altra, permette di prendere in considerazione la costituzione di archivi nuovi, conservati su schede perforate, che rimandano non soltanto a un nuovo sistema di classificazione, ma soprattutto a una critica documentaria diversa da quella del secolo XIX. […] La coerenza [tra i dati] è assicurata […] da un minimo di formalizzazione del documento, in modo che si possano ritrovare, per un lungo periodo di tempo e per ogni unità di tempo, gli stessi dati, nella stessa successione logica. Da questo punto di vista, l'utilizzazione del calcolatore elettronico per lo storico non costituisce solo un immenso progresso pratico, per il risparmio di tempo, [ma] è anche una coazione teorica molto utile, nella misura in cui la formalizzazione di una serie documentaria destinata a essere programmata obbliga fin dall'inizio lo storico a rinunciare alla sua ingenuità epistemologica, a costruire il suo oggetto di ricerca, a riflettere sulle sue ipotesi, e a passare dall'implicito all'esplicito»[56].

5. Conclusioni

Il concetto di *invented archive* rappresenta una sfida e allo stesso tempo un'opportunità per l'archivistica e la ricerca storica contemporanea, sebbene probabilmente non in egual misura per i due ambiti. Da un lato, la digitalizzazione e l'emergere di archivi inventati, nel contesto del web e delle tecnologie digitali, pongono questioni fondamentali sulla natura, sulla funzione e sul valore degli archivi tradizionali. Essi, infatti, non rispondono pienamente ai principi classici dell'archivistica, basati sulla relazione tra soggetto produttore, sedimentazione, uso e conservazione del documento. La loro creazione è guidata da motivazioni e scopi che variano dal desiderio di accesso aperto e democratizzazione della conoscenza, alla necessità di colmare lacune documentarie attraverso l'aggregazione di fonti disperse. Questa varietà di intenti, pur promuovendo l'accesso a informazioni storiche, rischia di sminuire l'importanza del vincolo archivistico e della coerenza interna che caratterizzano gli archivi tradizionali. In altre parole, un archivio inventato riflette più l'intenzione di chi lo crea che non la realtà documentaria del soggetto produttore, ponendo quindi un problema epistemologico per la disciplina archivistica. D'altra parte, gli *invented archives* offrono alla ricerca storica grandi prospettive e possibilità. Il carattere tematico e la struttura non gerarchica permettono di ripensare l'organizzazione e l'analisi dei dati storici in modi che sfuggono alle convenzioni archivistiche tradizionali. L'infrastruttura digitale del web facilita l'aggregazione e la consultazione di documenti provenienti da *repository* diversi, superando i limiti fisici e logistici degli archivi tradizionali. Inoltre, la possibilità di collegare documenti attraverso hyperlink e di manipolare digitalmente i dati, crea un ambiente di ricerca dinamico in cui le fonti possono essere continuamente riorganizzate e reinterpretate. Questo approccio dinamico ricorda, in parte, la metodologia dell'*histoire sérielle* degli anni Cinquanta del Novecento, che privilegiava l'analisi quantitativa e la serialità dei dati per spiegare i fenomeni storici. Gli archivi inventati possono, infatti, essere visti come una moderna evoluzione di quella metodologia in cui la serialità si integra con la flessibilità della struttura digitale, consentendo una più ampia gamma di interpretazioni e analisi storiche.

[56] Ibidem, p. 9–10.

MODUS

Tuttavia, mentre gli archivi inventati ampliano le possibilità di ricerca storica, pongono anche rischi significativi. La mancanza di criteri standardizzati di classificazione, descrizione e conservazione dei documenti nei contesti digitali può compromettere l'affidabilità delle fonti e la qualità della ricerca. Il fatto che l'accesso ai documenti digitali possa essere mediato solo da motori di ricerca o che la localizzazione dei dati dipenda da URL specifici, evidenzia una fragilità intrinseca nella loro archiviazione. L'assenza di un inventario formalizzato e di un sistema coerente di classificazione compromette la capacità degli archivi inventati di servire come depositi affidabili di conoscenza storica. Ciò solleva ulteriori interrogativi sulla loro sostenibilità a lungo termine e sull'integrità dei dati che custodiscono.

Nel contesto dell'archiviazione del web e delle sue sfide associate, emerge un altro punto cruciale: la necessità di un approccio ibrido che possa integrare i benefici della digitalizzazione con i principi tradizionali dell'archivistica. La creazione di archivi digitali non dovrebbe limitarsi alla mera riproduzione di documenti cartacei in formato digitale, ma dovrebbe impegnarsi in una riorganizzazione dei dati che rispetti sia la logica dei sistemi di archiviazione tradizionali sia le nuove esigenze di accesso e interazione poste dagli ambienti digitali. Strumenti come Internet Archive, pur rappresentando un tentativo di conservare l'eredità digitale, mostrano limiti evidenti in termini di classificazione, accessibilità e conservazione a lungo termine. L'assenza di un soggetto produttore chiaro e la mancanza di una struttura coerente di collocazione archivistica evidenziano le difficoltà nel considerare tali progetti come veri e propri archivi in senso tradizionale.

Infine, se è vero che gli *invented archives* rispondono a una domanda crescente di accesso immediato e illimitato a risorse storiche, è altrettanto vero che questo nuovo paradigma di archiviazione richiede una riflessione più profonda e critica sul ruolo degli archivi nell'era digitale. L'archivistica deve, quindi, affrontare una duplice sfida: riconoscere e adattarsi ai cambiamenti tecnologici e metodologici portati dalla digitalizzazione e dalla creazione di archivi inventati, e al tempo stesso mantenere e riaffermare i propri principi fondamentali, i quali garantiscono l'integrità, la contestualizzazione e la continuità delle fonti documentarie. È essenziale che archivisti, storici e *computer scientists* collaborino per sviluppare metodologie che possano armonizzare questi due mondi apparentemente contrastanti, al fine di preservare l'affidabilità delle fonti storiche e promuovere una ricerca sia rigorosa che innovativa.

In questo senso, la definizione stessa di "archivio" che ha da sempre rappresentato una sfida per i teorici dell'archivistica che, così come descritto e ricostruito da Gianni Penzo Doria[57], si sono cimentati nella formulazione del suo enunciato, necessita di un aggiornamento che rifletta la nuova realtà della documentazione digitale, della struttura dei "luoghi" (i server e i supporti di connessione e memorizzazione, il ruolo delle aziende che forniscono gli spazi e i vari servizi cloud – accettando il rischio di *vendor lock-in*, ossia la creazione di un rapporto di dipendenza col fornitore), del ruolo dei *web users* (che, in un certo qual modo, lasciano "dati" sui server e nei vari siti internet al momento

[57] Gianni Penzo Doria, *A new archives definition*, «JLIS.It», 2 (2022), n. 13, p. 156–173.

MODUS

della consultazione) e che tenga conto delle specifiche di tali complessi documentari anche sotto l'ottica della *computer science*, pur rimanendo ancorati ai fondamenti teorici e metodologici della tradizione umanistica.

This essay explores the phenomenon of "invented archives" and its implications for archival science and historical research in the digital age. These new types of archives, often thematic and non-traditional, emerge as responses to digitization and the need for open access, challenging the classical principles of archival science based on the relationship between the record creator, accumulation, and documentary preservation. On another front, the essay delves into the theoretical origins of "invented archives," analyzing how these documentary collections draw inspiration from the methodologies of histoire sérielle developed in the 1950s, where seriality and the quantitative analysis of historical data were central to the construction of a new historiography. With the spread of digital technologies, these methodologies have evolved, leading to the creation of documentary complexes that do not follow traditional archival hierarchies but focus on flexibility and the reorganization of data.

La ricerca si è avvalsa di un finanziamento dell'Unione europea – Next Generation EU – missione 4, componente 2, investimento 1.1, nell'ambito del programma PRIN-PNRR. Il titolo del progetto è: A Database on the Slave Trade between the Mediterranean and the Atlantic (15th-16th centuries).

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

Riferimenti bibliografici

Espen J. J. Aarseth. *Cybertext: Perspectives on Ergodic Literature*. Baltimore: Johns Hopkins University Press, 1997.

Syed E. Ahmed. *Perspectives on big data analysis methodologies and applications*, ed. by International Workshop on Perspectives on High-Dimensional Data Analysis. Providence (RI): American Math. Soc., 2014.

Stefano Allegrezza. *Towards a new archival economy: the development of the discipline in the transition from analogue to digital*. «JLIS.It», 1 (2017), n. 8, p. 114–126.

Gianluca Braschi. *Archivi inventati*. «*Medium*», 19 luglio 2024, <https://braschi.cc/archivi-inventati-9d67aba2e83a>.

Fernand Braudel. *Scritti sulla storia*. Milano: Bompiani, 2003.

Vannevar Bush. *As we may think*. «Atlantic», 1 (1945), n. 176, p. 112-124.

Ernst Cassirer. *The philosophy of symbolic forms*. London and New York: Routledge, 2021.

Pierre Chaunu. *Séville et l'Atlantique, 1504-1650 : Structures et conjoncture de l'Atlantique espagnol et hispano-américain (1504-1650). Tome II, volume 1 : La conjoncture (1504-1592)*. Paris: SEVPEN, 1959.

Pierre Chaunu. *Histoire quantitative ou histoire serielle*. Paris: Armand Colin, 1964.

Pierre Chaunu. *La durata, lo spazio e l'uomo nell'epoca moderna: la storia come scienza sociale*. Napoli: Liguori, 1983.

Giovanni Ciofalo — Silvia Leonzi. *Homo communicans. Una specie di/in evoluzione*. Roma: Armando Editore, 2013.

Alberto Clerici — Maurizio De Pra. *Informatica e web*. Milano: EGEA, 2012.

Leo Corry. *David Hilbert and the Axiomatization of Physics (1898–1918): From Grundlagen der Geometrie to Grundlagen der Physik*. Londra: Kluwer Academic Publishers, 2004.

Robert Darcy — Richard C. Rohrs. *A guide to quantitative history*. Westport (Conn.): Praeger, 1995.

Charles M. Dollar — Richard J. Jensen. *Historian's guide to statistics; quantitative analysis and historical research*. Huntington (N.Y.): R.E. Krieger, 1974.

Gianni Penzo Doria. *A new archives definition*. «JLIS.It», 2 (2022), n. 13, p. 156–173.

MODUS

Joris van Eijnatten — Toine Pieters — Jaap Verheul. *Big Data for Global History: The Transformative Promise of Digital Humanities*. «BMGN - Low Countries Historical Review», 4 (2013), n. 128, p. 55–77.

Roberto Franzosi. *A third road to the past? Historical scholarship in the age of big data*. «Historical Methods: A Journal of Quantitative and Interdisciplinary History», 4 (2017), n. 50, p. 227–244.

Gottlob Frege. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry Into the Concept of Number*. Evanston: Northwestern University Press, 1980.

Gottlob Frege. *Senso, funzione e concetto. Scritti filosofici 1891-1897*. Bari: Laterza, 2007.

Gottlob Frege. *L'alfabeto del pensiero*. Roma: Lit Edizioni, 2021.

Gottlob Frege. *On Sense and Reference*. Londra: Routledge, 2011.

François Furet. *Sur quelques problèmes posés par le développement de l'histoire quantitative*. «Social Science Information», 1 (1968), n. 7, p. 71–82.

François Furet. *L'histoire quantitative et la construction du fait historique*. «Annales. Histoire, Sciences Sociales», 1 (1971), n. 26, p. 63–75.

François Furet. *Il quantitativo in Storia*. In: *Fare storia. Temi e metodi della nuova storiografia*, a cura di J. Le Goff, P. Nora. Torino: Einaudi, 1981.

Kurt Gödel. *The Consistency of the Axiom of Choice and of the Generalized Continuum-hypothesis with the Axioms of Set Theory*. Princeton (N.J.): Princeton University Press, 1940.

Kurt Gödel. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. New York: Dover, 1992.

Teofilo Gonzalez — Jorge Diaz-Herrera — Allen Tucker. *Computing Handbook, Third Edition: Computer Science and Software Engineering*. Boca Raton-London-New York: CRC Press, 2014.

Shawn Graham — Ian Milligan — Scott Weingart. *Exploring Big Historical Data. The Historian's Macroscope*, Singapore: World Scientific Publishing, 2015.

Maria Rosaria Guercio. *Archivistica informatica: i documenti in ambiente digitale*. Roma: Carocci, 2019.

David Hilbert. *Grundzüge der Theoretischen Logik*. Berlin: Springer, 1972.

David Hilbert. *Mathematical Problems. Lecture delivered before the International Congress of Mathematicians at Paris in 1900*, Lancaster & New York: Macmillan Company, 2021.

David Hilbert — Stephan Cohn-Vossen. *Geometry and the Imagination*. Providence (RI): AMS Chelsea publishing, 1999.

Oscar Itzcovich. *Lo storico e il database*. «Quaderni storici», 70 (1989), n. 24, p. 321–325.

Oscar Itzcovich. *L'uso del calcolatore in storiografia*. Milano: FrancoAngeli, 1993.

Oscar Itzcovich. *Dal mainframe al personal, il computer nella storiografia quantitativa*. In: *Storia e Computer: alla ricerca del passato con l'informatica*, a cura di S. Soldani e L. Tomassini. Milano: Mondadori, 1996, p. 30-47.

Klaus Bruhn Jensen. *Metamedium*. *The International Encyclopedia of Communication Theory and Philosophy*, Chichester: Wilet-Blackwell, 2016, p. 1–4.

Takeshi Kojima — Shixiang Shane Gu — Machel Reid — Yutaka Matsuo — Yusuke Iwasawa. *Large Language Models are Zero-Shot Reasoners*. arXiv, 2023, <https://arxiv.org/abs/2205.11916>.

Richard J. Lane. *The big humanities: digital humanities/digital laboratories*. New York: Routledge, 2017.

John McCarthy — Nathaniel Rochester — Marvin L. Minsky — Claude E. Shannon. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955*. <https://upload.wikimedia.org/wikipedia/commons/f/f4/A_Proposal_for_the_Dartmouth_Summer_Research_Project_on_Artificial_Intelligence%2C_by_John_McCarthy_et_al%2C_1955.pdf>.

Ministero per i Beni Culturali e Ambientali. Ufficio Centrale per i Beni Archivistici — Archivio di Stato di Torino — Regione Piemonte. Assessorato alla Cultura — CSI Piemonte — Istituto Bancario S. Paolo di Torino. *Informatica e Archivi. Atti del Convegno. Torino 17-19 giugno 1985*. Roma: Istituto Poligrafico e Zecca dello Stato ; Libreria dello Stato, 1986.

Ted H. Nelson. *Complex information processing: a file structure for the complex, the changing and the indeterminate.* In: *ACM '65: Proceedings of the 1965 20th national conference.* New York: Association for Computing Machinery, 1965, p. 84-100.

Tito Orlandi. *Informatica umanistica*. Roma: La Nuova Italia Scientifica, 1990.

Discipline umanistiche e informatica: il problema della formalizzazione, a cura di Tito Orlandi. Roma: Accademia Nazionale dei Lincei, 1997.

Tito Orlandi. *Formal Methods in the Humanities and their Teaching*. Presentato al *The future of the humanities in the digital age,* Bergen, 1998.

Edwin D. Reilly — Anthony Ralston — David Hemmendinger. *Encyclopedia of Computer Science*. London: Nature Publishing Group, 2000.

Francisco Rodriguez-Consuegra. *Kurt Gödel: Unpublished Philosophical Essays.* Basel-Boston-Berlin: Birkhäuser Verlag, 1995.

Joan Roselló. *Hilbert, Göttingen and the Development of Modern Mathematics.* Newcastle upon Tyne: Cambridge Scholars Publishing, 2019.

MODUS

Roy Rosenzweig. *Brave New World or Blind Alley? American History on the World Wide Web*. «Journal Of American History», 1 (1997), n. 84, p. 132–155.

Roy Rosenzweig. *The Road to Xanadu: Public and Private Pathways on the History Web*. «Journal of American History», 2 (2001), n. 88, p. 548–579.

Robert Rowland. *L'Informatica e il mestiere di storico*. «Quaderni storici», 3 (1991), n. 26, p. 693–720.

Robert Rowland — Renzo Derosas. *Premessa*. «Quaderni Storici», 78 (1991), n. 3., p. 687–691.

Enrica Salvatori. *Digital (Public) History: la nuova strada di una antica disciplina*. «RiMe. Rivista dell'Istituto di Storia dell'Europa Mediterranea», (2017), p. 57–94.

Thomas A. Sebeok. *Encyclopedic Dictionary of Semiotics*. Berlin-Boston: De Gruyter Mouton, 1986.

Salvatore Spina. *Datificazione delle fonti storiche per la digital history delle pandemie*. «Umanistica Digitale», (2021), n. 10, p. 1-18.

Salvatore Spina. *Digital History. Metodologie informatiche per la ricerca storica*. Napoli: Edizioni Scientifiche Italiane, 2022.

Salvatore Spina. *Homo-Loggatus. The anthropological condition of historians in the digital world*. «Journal of Mathematical Techniques and Computational Mathematics», 10 (2023), n. 2, p. 431–437.

Salvatore Spina. *"Homo-loggatus": uomo e macchine nella nicchia ecologica digitale*. «Informatica Umanistica e Cultura Digitale: il blog dell'AIUCD», 19 luglio 2023, <https://infouma.hypotheses.org/2224>.

Salvatore Spina. *Digitality as a longue durée historical phenomenon*. «Umanistica Digitale», (2024), n. 18, p. 1–25.

Dennis A. Trinkle — Scott A. Merriman. *The history highway 2000, a guide to Internet resources*. Armonk (N.Y.): M.E. Sharpe, 2000.

Alan Mathison Turing. *Computing machinery and intelligence*. «Mind», 236 (1950), n. LIX, p. 433–460.

Federico Valacchi. *Se l'archivio è artificiale: Verso uno ius archivi partecipativo?*. «AIDAinformazioni», (2023), n. 1–2, p. 153–170.

Stefano Vitali. *Passato digitale. Le fonti dello storico nell'era del computer*. Milano: Bruno Mondadori, 2004.

Alfred North Whitehead — Bertrand Russell — Kurt Gödel. *Principia mathematica: Vorwort und Einleitungen*. Frankfurt am Main: Suhrkamp, 1986.

Zhiwei Xu — Jialin Zhang. *Computational Thinking: A Perspective on Computer Science*. Singapore: Springer Nature, 2022.

MODUS

Eventi e segnalazioni

Biblioteche e trasformazione digitale: un percorso formativo e-learning a cura della Scuola nazionale del patrimonio e delle attività culturali e dell'Associazione Italiana Biblioteche

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00152

Giovanni Bergamin – *Associazione Italiana Biblioteche (AIB)*
Gaia Caltavituro — **Marzia Piccininno** — **Andrea Razza**
Scuola nazionale del patrimonio e delle attività culturali

Nel corso dei secoli, le biblioteche hanno attraversato profonde trasformazioni, evolvendosi da luoghi di conservazione della scrittura a centri dinamici di accesso e diffusione della conoscenza. Oggi, nel pieno della rivoluzione digitale, si trovano a vivere un nuovo e profondo cambiamento: la diffusione pervasiva delle tecnologie digitali ha infatti determinato una trasformazione radicale e irreversibile nella società, nell'economia e nella cultura. In questo scenario, la capacità di gestire, interpretare e rendere accessibili grandi quantità di dati e contenuti digitali è divenuta non solo una competenza utile, ma una condizione essenziale per migliorare la qualità della vita e garantire l'inclusione culturale e informativa.

Questa rivoluzione interessa da vicino le biblioteche, chiamate a confrontarsi con nuove forme di testualità e nuovi modi di memorizzare, comunicare e condividere la conoscenza. Cambiano i modi di trasmettere il sapere, si diversificano le aspettative degli utenti, si moltiplicano le modalità di accesso ai contenuti. In questo contesto, le biblioteche – insieme a musei e archivi – continuano a rappresentare le fondamenta della "memoria registrata" della collettività, configurandosi come spazi essenziali per la costruzione di un sapere condiviso e per la promozione delle "conversazioni" culturali delle comunità.

Per affrontare questa transizione, le biblioteche si stanno rinnovando sul piano strutturale e professionale. Gli ambienti bibliotecari si configurano oggi come spazi ibridi, in cui le collezioni cartacee convivono con risorse digitali sempre più dinamiche, in grado di rispondere in modo mirato e personalizzato ai bisogni informativi, educativi e culturali delle persone. In questo scenario, la formazione continua e l'aggiornamento professionale dei bibliotecari diventano strumenti indispensabili per affrontare con consapevolezza e competenza le sfide dell'era digitale.

Investire nello sviluppo delle competenze significa, dunque, non solo garantire l'adattamento delle biblioteche al cambiamento in atto, ma soprattutto renderle protagoniste

MODUS

attive della trasformazione, capaci di guidare l'innovazione, favorire l'accesso inclusivo alla conoscenza e rafforzare il loro ruolo di presidi culturali centrali nelle società contemporanee.

Dicolab. Cultura al digitale

La centralità delle persone e dello sviluppo delle loro competenze è uno dei pilastri del Piano nazionale di digitalizzazione del patrimonio culturale (PND)[1], varato nel 2022 dall'Istituto centrale per la digitalizzazione del patrimonio culturale (ICDP) del Ministero della cultura nell'ambito del PNRR Cultura 4.0, Investimento 1.1 "Strategie e piattaforme digitali per il patrimonio culturale". Si tratta di uno strumento di pianificazione strategica finalizzato allo sviluppo di modelli, processi, metodi e regole per innescare una vera trasformazione dell'ecosistema, guidare le azioni dei soggetti che contribuiscono alla digitalizzazione, gestire le risorse digitali, aggiornare e riqualificare gli operatori del settore per mettere in atto quel cambiamento radicale necessario per far fronte alla transizione digitale.

Dicolab. Cultura al digitale[2] è il programma di formazione realizzato dalla Scuola nazionale del patrimonio e delle attività culturali[3] in attuazione del PND, un'iniziativa significativa per portata (con l'obiettivo di erogare almeno 40.000 *open badge* - certificazioni digitali delle conoscenze e delle abilità ottenute, riconosciute a livello europeo - entro giugno 2026), durata (quattro anni, dal 2022 al 2026) e pubblico diversificato a cui si rivolge: dipendenti del Ministero della cultura e di altre pubbliche amministrazioni, dipendenti di imprese, liberi professionisti, laureandi, specializzandi e dottorandi in materie attinenti all'ambito del patrimonio culturale digitale.

Gli obiettivi principali del programma sono la costituzione di un linguaggio condiviso sui temi portanti del patrimonio culturale digitale e l'aggiornamento delle competenze di base e specialistiche degli operatori di questo settore a livello sia di istituzioni, sia di singoli professionisti.

A novembre 2025 il catalogo di Dicolab, accessibile tramite la piattaforma di formazione a distanza della Scuola[4], conta un nutrito numero di corsi attivi in vari formati (MOOC, cicli di webinar, podcast) che affrontano, con un approccio globale, tutti i temi del sapere digitale: dagli aspetti trasversali, quali la comunicazione, il *prompting*, la gestione dei dati e la loro sicurezza, a quelli più specifici e di dominio come il *web archiving*, le regole di catalogazione, o l'uso del Building information Modeling (BIM) nei contesti del patrimonio culturale. A questi si affiancano laboratori in presenza per approfondire e sperimentare quanto appreso a livello teorico, e la programmazione di dieci hub sul territorio nazionale che integrano i corsi online con una proposta di

[1] https://digitallibrary.cultura.gov.it/.
[2] https://dicolab.it/.
[3] La Scuola nazionale del patrimonio e delle attività culturali, <https://www.fondazionescuolapatrimonio.it/>, è un istituto internazionale per la formazione, la ricerca e gli studi avanzati nell'ambito delle competenze del Ministero della cultura, socio fondatore e finanziatore.
[4] https://fad.fondazionescuolapatrimonio.it/.

azioni formative in presenza[5], soprattutto di natura laboratoriale, e che mirano all'attivazione di processi di ascolto, al coinvolgimento del sistema culturale locale e allo sviluppo di alleanze progettuali.
La progettazione dei percorsi Dicolab ha attraversato diverse fasi e modalità: se fino a tutto il 2023 la Scuola ha puntato su temi trasversali ai vari settori del patrimonio culturale per costruire un linguaggio e conoscenze di base, successivamente ha circoscritto e approfondito quanto d'interesse per specifiche categorie di professionisti, anche e soprattutto attraverso un processo di co-creazione con istituti e reti.
È in tale contesto che si inserisce la collaborazione tra la Scuola nazionale del patrimonio e delle attività culturali e AIB - Associazione Italiana Biblioteche le quali, dopo aver valutato i fabbisogni formativi della comunità, hanno varato il ciclo di formazione a distanza "Biblioteche e trasformazione digitale". I destinatari sono professionisti e professioniste che lavorano con e per le biblioteche, coloro che operano negli istituti culturali e nel mondo MAB (Musei, Archivi e Biblioteche), e chiunque intenda approfondire l'ecosistema delle biblioteche digitali o diventare professionista del settore.

Biblioteche e trasformazione digitale

L'idea di un corso formativo sul rapporto tra biblioteche e trasformazione digitale nasce alla fine del 2022, all'interno dell'AIB, su iniziativa dell'allora presidente Rosa Maiello, a seguito dell'indagine sui fabbisogni formativi degli associati che evidenziava le "biblioteche digitali" tra i temi più richiesti. Il lavoro preparatorio si è fondato sulla consapevolezza che la trasformazione digitale rappresenti un nodo cruciale per le biblioteche, connettendo ambiti essenziali della società: il mondo del libro e della lettura, il patrimonio culturale, l'informazione, la comunicazione, l'istruzione, l'apprendimento permanente e la ricerca scientifica. Per questo, si è scelto di esplorare i diversi aspetti del rapporto tra biblioteche e trasformazione digitale, adottando una prospettiva operativa, nell'ottica di costruire una sorta di "manuale di sopravvivenza". A tal fine, la formula di un corso "a più voci" si è rivelata naturale, coinvolgendo docenti esperti dei temi trattati. Una tappa fondamentale è stata, nel 2023, l'avvio del dialogo tra AIB e la Scuola Nazionale del patrimonio e delle attività culturali. Questo confronto, proseguito sotto la guida della nuova presidente AIB Laura Ballestra (da giugno 2023), ha portato nel 2024 alla definizione del progetto Biblioteche e trasformazione digitale: un ciclo formativo erogato a distanza e articolato in due parti, "Biblioteche digitali" e "Biblioteche digitali 2 – Servizi alla cittadinanza", ciascuna composta di nove corsi[6] per fornire un aggiornamento professionale sui temi e sugli strumenti che fondano le biblioteche digitali.
Gli argomenti dei diciotto corsi sono articolati intorno a due temi principali: definire gli strumenti di base necessari per comprendere i principi fondamentali delle biblioteche digitali, approfondendone le soluzioni organizzative, le tecnologie, le procedure, gli stan-

[5] https://dicolab.it/hub/.
[6] I corsi del ciclo formativo sono stati pubblicati con cadenza bisettimanale sulla piattaforma FAD della Scuola e sono fruibili, previa registrazione, ai seguenti link: <https://fad.fondazionescuolapatrimonio.it/enrol/index.php?id=613> e <https://fad.fondazionescuolapatrimonio.it/enrol/index.php?id=707>.

dard, i vincoli normativi e istituzionali necessari a garantire l'accessibilità permanente e l'interoperabilità, e analizzare i servizi digitali per le comunità di riferimento, esplorando le potenzialità offerte dalle tecnologie di Intelligenza Artificiale, dei social media, del web e dei *discovery tools* e approfondendo i temi di *reference* digitale, *information* e *media literacy*, e scienza aperta.

Ogni corso si compone di un webinar della durata di un'ora e di materiali di approfondimento ai temi trattati, al termine della cui fruizione ogni utente ha la possibilità di sostenere un test di valutazione per l'ottenimento dell'*open badge*. Anche l'AIB riconosce ai suoi associati e a coloro che intendono associarsi 1 credito formativo (corrispondente a 25 ore di carico di lavoro) per ciascuno dei due percorsi. I crediti maturati da coloro che intendano ottenere l'attestazione AIB relativa al possesso dei requisiti di qualificazione professionale[7] contribuiranno a costituire il monte ore formativo (100 ore) richiesto dal Regolamento di iscrizione all'Associazione. I crediti maturati dagli associati AIB che desiderino richiedere il rinnovo dell'attestazione contribuiranno a costituire il numero di crediti (1 credito annuo) per i 5 anni successivi la prima attestazione o il rinnovo precedente (così come previsto dalla direttiva AIB sulla formazione continua).

Biblioteche digitali: le basi dell'ecosistema

La prima parte del percorso formativo fornisce un quadro di riferimento sull'ecosistema delle biblioteche digitali, intese come infrastrutture cruciali per l'accesso alle conoscenze, alle risorse e ai servizi digitali[8] (Fig. 1). Esplorandone i fondamenti, le tecnologie, le potenzialità e le criticità si approfondiscono i temi legati all'organizzazione, allo sviluppo e alla gestione delle biblioteche digitali, evidenziando i concetti chiave utili per orientarsi all'interno di un contesto tanto ricco quanto frammentato.

Si affrontano le sfide legate al potenziale dei dati bibliografici per l'analisi computazionale, illustrando strategie per garantire l'inclusione e mitigare eventuali *bias* nei processi di metadatazione. Sono definiti il contesto normativo delle biblioteche digitali, trattando questioni legate all'accesso, alla conservazione, all'uso e alla tutela dei dati del patrimonio culturale, e offrendo strumenti pratici per navigare un quadro giuridico complesso. In seguito, sono trattati il tema dell'identità digitale, dell'accesso ai servizi digitali, dell'autenticazione e della sicurezza, illustrando gli strumenti adottati a livello nazionale ed europeo.

[7] L'attestazione del possesso dei requisiti di qualificazione professionale viene rilasciata ai sensi dell'art. 7, comma 1 della Legge 14 gennaio 2013 n. 4.

[8] I nove corsi della prima parte del ciclo formativo sono i seguenti: 1. *Introduzione alle biblioteche digitali* (Maurizio Messina - già Direttore della Biblioteca Nazionale Marciana); 2. *Gestire collezioni bibliografiche digitali* (Rossana Morriello - Ricercatrice di Scienze biblioteconomiche, Università degli studi di Firenze); 3. *Accesso e riuso delle risorse digitali: aspetti giuridici* (Rosa Maiello - Direttrice Biblioteca, Università Parthenope di Napoli); 4. *Identità digitale e accesso ai servizi* (Valdo Pasqui - già Responsabile Coordinamento Tecnico Applicativi - SIAF); 5. *Gestione delle entità digitali* (Carlo Camillo Pellizzari di San Girolamo - Dottorando in Scienze dell'antichità, Scuola Normale Superiore); 6. *Conservazione e sicurezza dei beni culturali digitali* (Emanuele Bellini – Ricercatore di Database and Digital Humanities, Università degli Studi Roma Tre); 7. *Progettare la digitalizzazione delle raccolte* (Gianfranco Crupi - Professore di Storia del libro antico e Biblioteconomia, Sapienza - Università di Roma); 8. *Classificazioni bibliografiche per il digitale* (Piero Cavaleri - Editor Web Dewey italiana); 9. *Web semantico e Linked data* (Lucia Sardo – Ricercatrice di Storia del libro e biblioteconomia e Catalogazione e classificazione bibliografica, Alma Mater Studiorum, Università di Bologna).

Figura 1. *Copertina della prima parte del percorso formativo pubblicato sulla piattaforma FAD della Scuola nazionale del patrimonio e delle attività culturali*

Sono quindi approfondite le caratteristiche del software libero usato da Wikidata, Wikibase, esaminando in particolare il modo in cui i dati vengono strutturati e come possono essere interrogati. Si esplorano poi i temi legati alla conservazione e alla sicurezza delle risorse digitali culturali, analizzando metodologie e tecnologie per proteggerle da minacce fisiche e informatiche. Viene definito il processo di progettazione delle collezioni digitali, indagando formati, linguaggi e tecnologie utili alla loro creazione e gestione. È esaminato il ruolo delle classificazioni bibliografiche nel contesto delle risorse digitali e il loro adattamento alle nuove esigenze.
Infine, sono approfondite le caratteristiche e le tecnologie del web semantico, osservando come queste possano essere utilizzate per valorizzare il patrimonio culturale attraverso piattaforme collaborative e buone pratiche.

Biblioteche digitali 2 - Servizi alla cittadinanza

La seconda parte del ciclo formativo si propone di esplorare i servizi digitali delle biblioteche e il ruolo che ricoprono nel soddisfare le esigenze delle comunità di riferimento[9] (Fig. 2).

Figura 2. *Copertina della seconda parte del percorso formativo pubblicato sulla piattaforma FAD della Scuola nazionale del patrimonio e delle attività culturali*

Partendo da un'introduzione sulle potenzialità offerte dall'Intelligenza Artificiale, si esaminano le sue applicazioni pratiche e le criticità nel contesto bibliotecario. Si approfondisce il valore di web e *social media archiving* per la conservazione del patrimonio digitale e della memoria collettiva. Si esaminano le strategie di valutazione e le competenze necessarie per garantire l'affidabilità della ricerca e del recupero della documentazione di fonte pubblica digitale. Si analizzano le tecnologie e le pratiche che supportano l'accesso libero alla conoscenza e alla scienza aperta, e le strategie per sostenere un dialogo efficace con la comunità attraverso siti web e social media. Si approfondiscono le modalità attraverso cui i bibliotecari, sia con il servizio di *reference* digitale che attraverso corsi

[9] La seconda parte del ciclo formativo è composta dei seguenti corsi: 1. *Biblioteche e IA. Introduzione e applicazioni* (Lorenzo Gobbo - Addetto ai servizi digitali presso le biblioteche, Università della Svizzera italiana); 2. *Introduzione al web e social media archiving* (Beatrice Cannelli - Ricercatrice, Bodleian Libraries University of Oxford); 3. *Identificare e valutare la documentazione di fonte pubblica digitale* (Lucia Antonelli - Funzionario, Biblioteca della Direzione centrale per le Autonomie. Albo nazionale dei Segretari comunali e provinciali, Ministero dell'Interno); 4. *Biblioteche digitali e scienza aperta* (Anna Maria Pastorini - Università di Genova e Coordinatrice del Gruppo di Studio sulle politiche dell'informazione - AIB); 5. *Information e media literacy. Educare alla competenza informativa in biblioteca* (Federica Viazzi - Bibliotecaria, Osservatorio AIB sull'information literacy e Biblioteca Biomedica – DAIRI Ospedale SS. Antonio e Biagio e Cesare Arrigo, Alessandria); 6. *La comunicazione in biblioteca: strumenti, strategie, visioni* (Cristina Bambini - Direttrice, Biblioteca Comunale di Cattolica); 7. *Comunicare la biblioteca attraverso il suo sito web* (Andrea Marchitelli - Regional Sales Manager EBSCO e Vice-Coordinatore AIB-WEB); 8. *Il servizio di reference digitale* (Francesca Papi - Coordinatrice Gestionale, Biblioteca Centrale "Leon Battista Alberti" Campus di Cesena - Alma Mater Studiorum, Università di Bologna); 9. *L'evoluzione del catalogo e i discovery tools* (Andrea Marchitelli - Regional Sales Manager EBSCO e Vice-Coordinatore AIB-WEB).

MODUS

e laboratori, aiutano i cittadini a sviluppare la loro competenza mediale e informativa, e a fare un uso critico e consapevole delle informazioni. Si esplora infine la storia e l'evoluzione del catalogo e dei *discovery tools* nel mondo contemporaneo.

I risultati

La necessità della realizzazione di un percorso formativo sui temi appena elencati e la grande richiesta da parte della comunità dei professionisti e degli appassionati di un continuo aggiornamento sulle nuove traiettorie intraprese dal mondo delle biblioteche è evidenziata dal successo ottenuto da questi corsi in termini di fruizione. I numeri che emergono dall'analisi delle utenze dipingono un quadro che vede 5.257 utenti iscritti al percorso o ai singoli corsi, e ben 14.370 *open badge* totali rilasciati[10], che fanno di "Biblioteche digitali" uno dei cicli formativi di maggior successo di *Dicolab. Cultura al digitale*.

L'importanza e la trasversalità degli argomenti approfonditi, e la possibilità di usufruire degli oggetti formativi anche in modalità on demand, poi, rappresentano un'opportunità per chiunque voglia esplorare la materia, e il costante aumento delle utenze dei corsi testimonia la buona riuscita del progetto e l'importanza della formazione continua e della collaborazione tra enti culturali, obiettivo focale della Scuola e di AIB.

[10] Dati aggiornati al 18 novembre 2025.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

Europeana Climate Action Community and the path to regenerative digital transformation

«DigItalia» 2-2025
DOI: 10.36181/digitalia-00153

Evangelia Paschalidou
International Hellenic University

Europeana is the EU initiative dedicated to supporting the digital transformation of the European cultural heritage sector[1]. Starting as a platform for collecting, preserving and sharing EU's digital cultural collections, Europeana's mission has evolved into fostering a more accessible, inclusive, and interactive space - the data space[2]. The Europeana ecosystem is built on a collaborative structure involving the Europeana Foundation[3], the Europeana Network Association (ENA)[4] and the Europeana Aggregators' Forum[5], achieving multi-level digital capacity building.

One of the key thematic communities within the ENA is the Climate Action Community (CAC)[6]. Founded in response to growing environmental concerns, the CAC brings together practitioners committed to embedding environmental sustainability in digital cultural heritage. It serves as a collaborative hub for collective learning and action, enabling professionals from across the sector to explore sustainable and regenerative alternatives to "digital as usual". Aligned with the goals of the European Green Deal[7] and the EU Work Plan for Culture (2023-2026)[8], the CAC has emerged as a driving force in addressing one of the most pressing challenges of our time: the environmental footprint of digital cultural heritage.

A community response: The Environmental Sustainability Practice Task Force

Within the Climate Action Community, the Environmental Sustainability Practice Task Force[9] was born in 2022 out of a need; we aimed to explore how digital preservation practices can be made more environmentally responsible by cultural heritage institutions (CHIs), however, we had no data of the real-life practices employed and how they

1 https://www.europeana.eu/it/about-us.
2 https://pro.europeana.eu/page/common-european-data-space-for-cultural-heritage.
3 https://pro.europeana.eu/about-us/foundation.
4 https://pro.europeana.eu/europeana-network-association.
5 https://pro.europeana.eu/page/aggregators.
6 https://pro.europeana.eu/page/climate-action-community.
7 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.
8 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022G1207(01).
9 https://pro.europeana.eu/project/environmental-sustainability-practice-task-force.

MODUS

are scoring in terms of sustainability. To investigate this issue in depth, the Task Force launched a sector-wide Sustainability Practice Survey in 2023, that gathered insights from 108 professionals across 24 EU Countries and six others. Libraries, academic and research institutions, museums, and archives were all represented.

The results are making up the report Regenerative Digital Transformation: Sustainable Pathways for Cultural Heritage Institutions – a comprehensive analysis of how environmental sustainability is (or isn't) embedded across the full digital preservation lifecycle[10]. To deepen the analysis, three case studies were conducted with the National Library of Finland, the International Museum network, and POLIN Museum in Warsaw. These in-depth interviews revealed promising practices, as well as the barriers that CHIs face when attempting to adopt more sustainable digital methods.

Key insights: where we stand and what needs to change

The report reveals a growing awareness of environmental responsibility in the sector: around 80% of CHIs acknowledge their responsibility. Yet only 42% of surveyed institutions have formal environmental strategies, and a mere 14% track the carbon footprint of their digital services.

Most notably, many organisations continue to treat digital preservation as a siloed, project-based task – lacking coordinated strategies for content selection, long-term storage, and data lifecycle management. In fact, 57% of institutions lack policies governing the deaccessioning or retirement of digital assets, contributing to the rise of so-called "dark data", that is data preserved by organisations but never accessed. Also, redundant data storage practices are widespread –86% of CHIs keep three or more backup copies of their digital assets, yet few assess the environmental cost of these practices.

Coming to hardware, a major blind spot is procurement. While 17% of respondents report adopting recycling initiatives, only 8% consider hardware repairability, and a strikingly low 3% include sustainability criteria when purchasing new equipment. Links greatly back to the environmental cost of ICT infrastructure, not just by emphasizing the energy consumption of EU data centers (up to 65 TWh of electricity in 2022), but mainly by highlighting the hidden impact of manufacturing digital devices, which often rely on rare minerals extracted under environmentally and socially exploitative conditions.

These are only a few of the insights that are calling for a paradigm shift in how CHIs think about sustainability. The report is taking that a step further and encourages CHIs to adopt a regenerative practice. That means to design their digital preservation processes in a way that – rather than focusing solely on reducing harm – to actively restore ecosystems, support communities, and promote equity.

[10] https://pro.europeana.eu/post/regenerative-digital-transformation-sustainable-pathways-for-cultural-heritage.

Our collective way forward

The Europeana Climate Action Community's work is a timely reminder that cultural heritage is not just about the past – it is also about the kind of future we want to shape, while facing a climate crisis. The digital preservation decisions made today will impact generations to come, and as stewards of memory, CHIs hold the power to lead this transformation, and this boils down to aligning their mission of safeguarding memory with a commitment to environmental justice.

The full report, including survey results, CHIs' recommendations, case studies, and policy recommendations, is available online at Europeana Pro[11].

The Climate Action Community invites all cultural professionals to join and be part of this ongoing, sector-wide effort toward a more sustainable – and regenerative – digital future.

[11] See the previous footnote.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2025.

MODUS

ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane
www.iccu.sbn.it

Copyright © ICCU - Roma

2025 - Istituto Poligrafico e Zecca dello Stato S.p.A. - Roma
Finito di stampare nel mese di dicembre 2025
presso lo stabilimento IPZS di Via Salaria, 691 Roma

MODUS